2-1-2022

Development of Guidelines for Collecting Transit Ridership Data

Hong Yang
Old Dominion University, hyang@odu.edu

Kun Xie
Old Dominion University, kxie@odu.edu

Sherif Ishak
Old Dominion University, sishak@odu.edu

Qingyu Ma
Old Dominion University

Yang Liu
Old Dominion University

Original Publication Citation

Hong, Y., Xie, K., Ishak, S., Ma, Q., & Yang, L. (2022). *Development of guidelines for collecting transit ridership data* (Publication No. FHWA/VTRC 22-R22). Virginia Transportation Research Council. http://www.virginiadot.org/vtrc/main/online_reports/pdf/22-r22.pdf

VTRC
Virginia Transportation
Research Council

We Bring Innovation to Transportation

Development of Guidelines for Collecting Transit Ridership Data

http://www.virginiadot.org/vtrc/main/online_reports/pdf/22-r22.pdf

HONG YANG, Ph.D.
Department of Computational Modeling, Simulation and Engineering

KUN XIE, Ph.D.
Department of Civil Engineering

SHERIF ISHAK, Ph.D.
Department of Civil Engineering

QINGYU MA, Ph.D.
Department of Computational Modeling, Simulation and Engineering

YANG LIU
Department of Computational Modeling, Simulation and Engineering

OLD DOMINION UNIVERSITY

Final Report VTRC 22-R22

VIRGINIA TRANSPORTATION RESEARCH COUNCIL
530 Edgemont Road, Charlottesville, VA 22903-2454
vtrc.virginiadot.org

Standard Title Page - Report on Federally Funded Project

1. Report No.: FHWA/VTRC 22-R22	2. Government Accession No.:	3. Recipient's Catalog No.:
4. Title and Subtitle: Development of Guidelines for Collecting Transit Ridership Data		5. Report Date: February 2022
		6. Performing Organization Code:
7. Author(s): Hong Yang, Ph.D., Kun Xie, Ph.D., Sherif Ishak, Ph.D., Qingyu Ma, Ph.D., and Yang Liu		8. Performing Organization Report No.: VTRC 22-R22
9. Performing Organization and Address: Virginia Transportation Research Council 530 Edgemont Road Charlottesville, VA 22903		10. Work Unit No. (TRAIS):
		11. Contract or Grant No.: 116838
12. Sponsoring Agencies' Name and Address: Virginia Department of Transportation 1401 E. Broad Street Richmond, VA 23219	Federal Highway Administration 400 North 8th Street, Room 750 Richmond, VA 23219-4825	13. Type of Report and Period Covered: Final Contract
		14. Sponsoring Agency Code:
15. Supplementary Notes: This is an SPR-B report.		

16. Abstract:

Transit ridership is a critical determinant for many transit applications such as operation optimizations and project prioritization under performance-based funding mechanisms. As a result, the quality of ridership data is of utmost importance to both transit administrative agencies and transit operators. Many transit operators in Virginia report their ridership data to the Department of Rail and Public Transportation (DRPT) and the National Transit Database (NTD). However, with no specific guidelines available to transit agencies in Virginia for collecting ridership data, the heterogeneous mixture of diverse data collection methods and technologies has often raised concerns about the consistency and quality of the reported data. This study investigated the ridership data collection practices adopted by transit agencies in Virginia and developed high-level guidelines to facilitate data collection with improved quality. Specifically, it examined the data collection practices discussed in the literature and those adopted by local transit agencies in Virginia. The research team surveyed 39 transit agencies to obtain a clear understanding of their current practices in data collection scope, technological solutions, sampling and estimation techniques, and data storage and reporting, among others. To evaluate the potential estimation errors based on sampled data, the researchers requested and obtained actual data from five transit agencies of different sizes in Virginia. Comparisons between selected data collection solutions were conducted, and the estimation errors were tested based on different sample data from these agencies. Based on the findings from literature review, surveys, and analysis of actual data, a set of high-level data collection guidelines was proposed. This study recommends that DRPT distribute the developed guidelines among transit agencies in Virginia to help facilitate improved data collection practices across Virginia. It is also recommended that DRPT require the submission of each agency's ridership data collection methods and correction (adjustment) procedures, in addition to the agency's reported ridership data.

17 Key Words: Transit Ridership; Transit Agency; Automated Passenger Counter; Farebox; Sampling; Data Collection		18. Distribution Statement: No restrictions. This document is available to the public through NTIS, Springfield, VA 22161.	
19. Security Classif. (of this report): Unclassified	20. Security Classif. (of this page): Unclassified	21. No. of Pages: 64	22. Price:

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

FINAL REPORT

DEVELOPMENT OF GUIDELINES FOR COLLECTING TRANSIT RIDERSHIP DATA

Hong Yang, Ph.D.
Department of Computational Modeling, Simulation and Engineering
Old Dominion University

Kun Xie, Ph.D.
Department of Civil Engineering
Old Dominion University

Sherif Ishak, Ph.D.
Department of Civil Engineering
Old Dominion University

Qingyu Ma, Ph.D.
Department of Computational Modeling, Simulation and Engineering
Old Dominion University

Yang Liu
Graduate Research Assistant
Department of Computational Modeling, Simulation and Engineering
Old Dominion University

VTRC Project Manager
Peter Ohlms, AICP, Virginia Transportation Research Council

In Cooperation with the U.S. Department of Transportation
Federal Highway Administration

Virginia Transportation Research Council
(A partnership of the Virginia Department of Transportation
and the University of Virginia since 1948)

Charlottesville, Virginia

February 2022
VTRC 22-R22

DISCLAIMER

The project that is the subject of this report was done under contract for the Virginia Department of Transportation, Virginia Transportation Research Council. The contents of this report reflect the views of the author(s), who is responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Virginia Department of Transportation, the Commonwealth Transportation Board, or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation. Any inclusion of manufacturer names, trade names, or trademarks is for identification purposes only and is not to be considered an endorsement.

Each contract report is peer reviewed and accepted for publication by staff of the Virginia Transportation Research Council with expertise in related technical areas. Final editing and proofreading of the report are performed by the contractor.

ABSTRACT

Transit ridership is a critical determinant for many transit applications such as operation optimizations and project prioritization under performance-based funding mechanisms. As a result, the quality of ridership data is of utmost importance to both transit administrative agencies and transit operators. Many transit operators in Virginia report their ridership data to the Department of Rail and Public Transportation (DRPT) and the National Transit Database (NTD). However, with no specific guidelines available to transit agencies in Virginia for collecting ridership data, the heterogeneous mixture of diverse data collection methods and technologies has often raised concerns about the consistency and quality of the reported data. This study investigated the ridership data collection practices adopted by transit agencies in Virginia and developed high-level guidelines to facilitate data collection with improved quality. Specifically, it examined the data collection practices discussed in the literature and those adopted by local transit agencies in Virginia. The research team surveyed 39 transit agencies to obtain a clear understanding of their current practices in data collection scope, technological solutions, sampling and estimation techniques, and data storage and reporting, among others. To evaluate the potential estimation errors based on sampled data, the researchers requested and obtained actual data from five transit agencies of different sizes in Virginia. Comparisons between selected data collection solutions were conducted, and the estimation errors were tested based on different sample data from these agencies. Based on the findings from a literature review, surveys, and analysis of actual data, a set of high-level data collection guidelines was proposed. This study recommends that DRPT distribute the developed guidelines among transit agencies in Virginia to help facilitate improved data collection practices across the Commonwealth. It is also recommended that DRPT require the submission of each agency's ridership data collection methods and correction (adjustment) procedures, in addition to the agency's reported ridership data.

TABLE OF CONTENTS

FINAL REPORT

DEVELOPMENT OF GUIDELINES FOR COLLECTING TRANSIT RIDERSHIP DATA

Hong Yang, Ph.D.
Department of Computational Modeling, Simulation and Engineering
Old Dominion University

Kun Xie, Ph.D.
Department of Civil Engineering
Old Dominion University

Sherif Ishak, Ph.D.
Department of Civil Engineering
Old Dominion University

Qingyu Ma, Ph.D.
Department of Computational Modeling, Simulation and Engineering
Old Dominion University

Yang Liu
Department of Computational Modeling, Simulation and Engineering
Old Dominion University

INTRODUCTION

Many transit agencies in the U.S. are embracing data-driven decision-making approaches in their system planning, operations, and reporting. Likewise, transit agencies in Virginia are also leveraging various data as crucial enablers for many applications (e.g., planning, operations, estimation for capital funding grant applications, etc.). Among various types of data, transit ridership data is one of the most important. For example, ridership is a key variable in the performance-based funding mechanism adopted by the Virginia Department of Rail and Public Transportation (DRPT) (DRPT, 2019), and it is used in Virginia's SMART SCALE project prioritization process for ranking and selecting proposed transit projects (VDOT, 2019). The Virginia Department of Transportation (VDOT) also leverages ridership data from transit operators for travel demand modeling (Cambridge Systematics, 2014). Accurate ridership data is of utmost importance to both transportation administrative agencies and transit operators because of these critical uses.

Currently, many Virginia transit agencies report ridership data several times a year to the DRPT and the National Transit Database (NTD) of the Federal Transit Administration (FTA). However, no specific guidelines for collecting ridership data are available to transit agencies. One of the more comprehensive studies on ridership data collection is a Transit Cooperative Research Program (TCRP) Synthesis published in 2008 (Boyle, 2008). Nevertheless, there is very limited discussion on the latest developments and practices in collecting ridership data, particularly in the Commonwealth of Virginia. The heterogeneous mixture of diverse data collection methods and technologies employed by Virginia's different transit agencies makes it very challenging for administrative agencies such as DRPT and FTA to oversee the data collection process and assess the quality of the reported ridership data. Unified guidance on collecting, sifting, validating, storing, and reporting ridership data is needed to facilitate the collection of high-quality data from transit agencies in an efficient and timely manner. In this project, the research team conducted an in-depth study of current ridership data practices among Virginia transit agencies and developed a set of guidelines to support Virginia transit agencies in collecting ridership data.

PURPOSE AND SCOPE

The overarching goal of this study was to develop a set of guidelines for collecting transit ridership data for Virginia transit agencies. The research team accomplished this goal by examining existing practices in ridership data collection at agencies in the U.S., especially in Virginia. The team identified the following key research questions in the ridership data collection process:

1. What are the current practices adopted by other U.S. transit agencies?
2. What are the practices of local transit agencies in Virginia?
3. What data quality issues are associated with current ridership data collection practices, and are there any ways to address the issues?

The scope of the study was limited to fixed-route buses and demand-response modes, including bus rapid transit (BRT) and vanpools in a limited way, as feasible. Ridership data collection practices specific to light rail, subway, rail, and ferry modes were outside the scope of this study, as was the development of approaches for estimating transit ridership data.

METHODS

The following main tasks were conducted to achieve the research objectives:

1. Conduct a literature review.
2. Survey transit agencies in Virginia.
3. Evaluate transit ridership data collection approaches.
4. Develop ridership data collection guidelines.

Conducting the Literature Review

To achieve a comprehensive review of existing research efforts, the research team conducted an extensive search for published work through the following sources: 1) Google Scholar; 2) Google search engine; and 3) Transport Research International Documentation (TRID). The primary search terms included combinations of "APC," "AFC," "farebox," "manual counting," "mobile app," "public transit ridership," "data collection," "vanpool," "demand response," "bus," and "bus rapid transit." Several key criteria used to filter more relevant studies included keywords related to the research topic, studies describing the U.S. context, publication dates after 2008, and access to full papers.

Surveying Transit Agencies in Virginia

The research team developed an online survey using the Qualtrics platform to elicit input from transit professionals (e.g., transit program managers) about their experiences in collecting transit ridership data. (Please see Appendix A for a sample survey.) Their contacts were identified through each agency's official website and/or with the help of DRPT and the Technical Review Panel (TRP). The survey questions covered data collection techniques, concerns, etc. The initial email invitation was sent to contacts in early November 2020, and phone calls were made to people on the contact list shortly after the initial survey invitation was sent to ask them to complete the survey. Reminder emails were sent to all respondents who had not completed the survey had not started it by mid-November and early December 2020. Assistance was also sought from the study's TRP to reach more transit agencies in the Commonwealth. Detailed survey questions can be found in Appendix A. The research team reached out to 54 agencies, and representatives of 39 transit agencies accessed the survey, although not all completed the survey. Twenty-two agencies listed in the 2019 DRPT Statewide Integrated Mobility Initiative were contacted, including 19 completed the survey and one partially responded.

Evaluating Transit Ridership Data Collection Approaches

Based on the survey results, the research team further contacted a subset of transit agencies across the Commonwealth and requested their historical transit ridership data to assess the quality of annual ridership estimation. The list of agencies contacted is shown in Table 1. These agencies were selected based on the scale and location. Finally, five agencies that have provided historical ridership data mentioned that they did not use sampled data for estimating annual ridership. Instead, their raw data collected through a specific approach (e.g., manual counts, farebox data, and APC data) were reported without adjustment based on samples. Due to privacy concerns or data availability issues, only system-level historical daily ridership data and/or daily ridership data from selected routes were provided to the research team. The subsequent analyses are based on the obtained data from five agencies.

Although trip-level data were not available, the sampling procedure provided by the NTD was adopted to evaluate the potential estimation error of annual ridership related to the use of sampled data. The original NTD sampling template for transit agencies to develop the minimum sample size for estimating metrics, including unlinked passenger trips (UPT) and passenger miles traveled (PMT), is shown in equation (1).

$$n = \left(\frac{Z_{0.95}}{\text{Precision}} \times \frac{std_{input}}{\mu_{input}} \right)^2 \times A = \left(\frac{1.96}{0.1} \times \frac{std_{input}}{\mu_{input}} \right)^2 \times 1.25 = \left(\frac{1.96}{0.1} \times CV_{input} \right)^2 \times 1.25 \qquad (1)$$

where n is the estimated sample size for the annual data; μ_{input}, std_{input}, and CV_{input} are the mean, standard deviation, and coefficient of variation of input data (e.g., historical data or data from prior samples), respectively; $Z_{0.95}$ is the value from the standard normal distribution for a 95% confidence level ($Z_{0.95}$ = 1.96); Precision denotes the degree to which the statistical estimates are precise, and A = 1.25 is a factor to provide a margin of safety.

Table 1. Contacted Transit Agencies for Historical Ridership Data

ID	Agency
1	Fairfax Connector
2	Winchester Transit
3	Radford Transit: by New River Valley Community Services
4	Williamsburg Area Transit Authority
5	City of Suffolk - Suffolk Transit
6	Loudoun County Transit
7	DASH (Alexandria Transit Company)
8	Potomac & Rappahannock Transportation Commission-OmniRide
9	Four County Transit
10	Hampton Roads Transit
11	Virginia Regional Transit
12	Jaunt, Inc.
13	Pulaski Area Transit
14	Danville Transit System
15	Mountain Empire Older Citizens, Inc.
16	Harrisonburg Transit

Because the trip-level samples were not available, the same sampling procedure was applied to the daily ridership data obtained from a system and/or route. The historical data for one fiscal year were used to determine the μ_{input}, std_{input}, and CV_{input}. The same factor A = 1.25 was used in calculation. As day-to-day service levels are typically very different, a random sample of days can be problematic and impractical in terms of implementing data collection. Instead, for each sample this study used a full week of data, which is typically 5 to 7 days, depending on whether the agency operates on weekends or other specific days. It should be noted that the sampling process can be further adjusted if agencies operate different service levels at different times (e.g., seasons) of the year. For the sake of simplicity, a full-week sampling plan was assessed. The minimum number of weeks is estimated by the research team based on the following steps:

- Step 1. Based on historical daily ridership data, use equation (1) to compute the estimated number of days n for sampling ridership data. CV_{input} is determined based on prior year's data.
- Step 2. Depending on the number of days Wd an agency operates services each week, estimate the number of weeks W to collect sample data as $W = n / Wd$. Round W up to the nearest integer.
- Step 3. Randomly sample W weeks within a year (indexed as week 1 to week 52 of the year) as the period for collecting sample data. Note that the agency may not operate a full week for some of these sampled weeks because of holidays or other reasons. So, the total number of days actually sampled n_s for these sampled weeks may be slightly less than $W \times Wd$.
- Step 4. Implement the sampling plan to obtain the sampled daily ridership in the target fiscal year. Assume the agency obtained sampled daily ridership records for n_s days as R_1, R_2, …, R_{ns}.

- Step 5. Estimate annual ridership: Suppose the agency operates N days in the target fiscal year. Its annual ridership can be estimated with equation (2), which scales the average of the samples obtained in Step 4.

$$\hat{R} = \frac{R_1 + R_2 + ... + R_{n_s}}{n_s} \times N \tag{2}$$

- Step 6. For evaluating the estimation error, assume that the actual daily ridership data in the target fiscal year is $R_1, R_2, ..., R_N$. The calculated percentage error γ based on the estimated ridership against the actual ridership of the target fiscal year is obtained by equation (3).

$$\gamma = \frac{\hat{R} - (R_1 + R_2 + ... + R_N)}{R_1 + R_2 + ... + R_N} \times 100 \tag{3}$$

- Step 7. Due to random sampling error, the sampled weeks may be different if the experiment is repeated. Thus, running Steps 3 through 6 with a different random seed can lead to different results. Assume each repeated experiment with a different random seed will result in the calculated percentage errors γ_k, where $k = 1, 2, \ldots, K$ and K is the total number of repeated experiments. One can show the distribution of the estimated error based on γ_k. For example, in one sampling experiment, an analyst may randomly pick Weeks 10, 14, 28, 35, and 47 as the data collection period. In another sampling experiment, the analyst might randomly pick Weeks 13, 26, 29, 38, and 43 as the data collection period. Thus, it is expected that the error of the estimated annual ridership based on the sample data from each of these two experiments will be different. Repeating such experiments will obtain a set of estimation errors and their distribution will be examined.

Developing Ridership Data Collection Guidelines

Based on the findings from the literature review on ridership data collection approaches, related practices, and ridership data issues, we customized the survey to gather information on ridership data collection among transit agencies in Virginia. Building upon the synthesized survey results and analysis of actual ridership data from a subset of transit agencies, we developed a set of guidelines on the implementation of data collection methods, sampling guidance, and data processing and reporting. They offer high-level guidance to DRPT and/or VDOT regarding how best to assist individual transit agencies via practices such as updated statewide reporting requirements.

RESULTS

Literature Review

Existing studies and practices related to public transit ridership data collection over the past few decades show that as technologies continue to advance, data collection approaches have evolved from traditional manual counting to a variety of automated data collection approaches. Current uses of collected transit ridership data include modeling transit ridership with contributing factors such as weather; predicting transit ridership trends for short- and long-term periods; and collecting, integrating, and validating transit ridership data from multiple data sources. This study explored the collection/integration and validation of ridership data via automatic passenger counting (APC), automated fare collection (AFC), electronic fareboxes, mobile apps, and manual counting.

Table 2Table is a summary of data collection technologies in identified studies. This aims to complement the studies on ridership data collection in Transit Cooperative Research Program (TCRP) Synthesis 77 (Boyle, 2008).

Table 2. Summary of Different Data Collection Technologies in the Literature

Data Collection Approach [a]	**# Studies**	**Pros**	**Cons**
Automatic passenger counting	26	Accurate Flexible due to diverse approaches such as infrared and cameras	Expensive to deploy Low coverage of buses equipped with automatic passenger counting Low coverage of agencies using automatic passenger counting compared with farebox
Manual	22	Relatively accurate Serves as benchmark	Introduce human error Time-consuming and expensive
Electronic farebox	19	Low cost Relatively accurate Widely used	Data collection error by transit drivers Need to combine with vehicle location information for metrics such as origin-destination
Automated fare collection	9	Precise in large cities High coverage rate	Limited by scale in small cities Potential counting errors due to trip transfers Need to combine with vehicle location information for metrics such as origin-destination
Mobile app	3	Convenient for collecting information	Users who do not use the app cannot be counted

[a] *Note: Some studies discussed more than one approach.*

As illustrated in Table 2, APC is one of the most studied approaches for ridership data collection, followed by manual counting and farebox methods. It should be noted that several state DOTs have reported fareboxes to be the most prevalent ridership data collection approach among transit agencies. For example, Kimley Horn and IBI Group (2019) summarized data collection approaches in Virginia. The level of deployment of technologies was divided into three categories: low (<50% of transit agencies), medium (50-75% of transit agencies), and high (>75% of transit agencies). APC was found to be at a low level of deployment in rural areas and medium in small urban areas/college towns and urban areas. On the other hand, farebox was found to be at a high level of deployment in all contexts. Thus, although a plurality of studies reviewed were research articles that used APC data, many existing transit operators rely on fareboxes, a more cost-effective method of collecting ridership data.

To summarize the findings shown in Table 2, the research team developed an interactive web interface to provide users with a visual representation of geographical locations and data collection methods reported in the literature review. The web interface is illustrated in Appendix B. The literature review included 29 identified references from both urban and rural areas. In 21 studies, manual ride checks on a fixed schedule were used to validate the collected ridership data. Fixed-route bus was the most common type of system examined in the studies (28), followed by vanpool (5), demand-response bus (2), and BRT (1). Detailed discussion regarding ridership data collection approaches and practices is presented next, followed by discussion of ridership data collection and validation issues.

Ridership Data Collection Approaches and Practices

APCs utilize various technologies to detect passenger boarding, such as infrared (IR) light beam cells, laser scanners, IR cameras, piezoelectric mats, and others. APCs are usually accompanied by an automated vehicle location (AVL) system. AVL systems collect and report locations of buses in operation. APCs using IR beam technology have a moderate cost but are prone to accuracy issues that need regular calibration and validation. For instance, Strathman et al. (2005) found that data collected using IR camera-based APCs required post-processing and validation to address over- and undercounting. On the other hand, APCs using video technology are expensive but tend to be more reliable than APCs using other technologies. For example, Monast et al. (2017) investigated the use of new processing algorithms to count passenger trips captured with pre-existing transit vehicle security cameras. After attempting multiple detector placements in multiple vehicle types with multiple camera configurations, the proposed algorithms demonstrated the setup was a cost-efficient way to count passenger trips repeatedly on the same vehicle. Several studies outside the U.S. have found that the accuracy of APC-collected data can be relatively high, e.g., 94% (Yang et al., 2010), 96% (García-Bunster and Torres-Torriti, 2008), and 97% (Yahiaoui et al., 2010). APCs using wireless device detection technologies use Bluetooth or Wi-Fi to count passengers' devices. For example, Dunlap et al. (2016) combined APCs that collected Bluetooth and Wi-Fi data with vehicle location data in Seattle to estimate riders' origins and destinations. However, it should be noted that such APCs are prone to issues of undercounting due to the low ratio of passengers who carry detectable wireless devices. Kostakos et al. (2013) found that only 12.8% of passengers carried devices that could be detected by Bluetooth-based APCs, though they may be more common now. Other approaches include APCs that monitor the air pressure of the ride suspension system of a transit vehicle, with a reported 97.6% accuracy ratio (Kotz et al., 2015).

Manual counting/manual ride check is a conventional approach and is required by the FTA for annual validation of ridership submissions based on APC/AVL systems and for NTD reporting. Manual counting can serve as the benchmark for periodic data validation and calibration. For example, Hampton Roads Transit (HRT, 2018) is currently using APC, farebox, and manual counting to collect ridership data. However, manual counting is prone to statistical error due to sampling variability. Shireman (2011) mentioned that drivers may take unusual measures on ride check days that can lead to potential ridership data errors. For example, drivers may try to give observers the impression that they are doing everything they can to stay on-time and miscount passenger boardings as a result.

Electronic fareboxes offer the benefit of easy deployment, relatively low cost, and the ability to continuously collect data. However, electronic farebox ridership data collection can be prone to several critical issues such as the inability to classify ridership by trip and bus stop (WAVE Transit, 2018). Also, certain fare types require bus operators to perform specific farebox functions that can introduce potential human operation errors. For example, Yang et al. (2015) examined potential errors such as duplicate, simultaneous, and outlier records. Results indicated that ridership and revenue may have been overestimated by up to 9.95% due to farebox data errors. In terms of trip and bus information, additional work is needed. For example, when WAVE Transit (2018) combined the farebox system and an AVL system to obtain stop-level ridership data, the accuracy of the farebox data was found to be inferior to that of APC-based technologies. Similarly, TCRP Synthesis 34 concluded that farebox counts were less accurate than conventional APC systems (Furth, 2000). The data accuracy of fareboxes was found to be 88% in one study (Peterson, 2013) and 91% in another (Oberli et al., 2010), as compared to 94% to 97% for APCs as already noted. The report of FCDOT (2015) noted that farebox data represented an average ridership for the month, and therefore, it was likely a more accurate figure than the one-day composite obtained through ride checks. However, fareboxes typically only provide a total ridership figure and do not provide the stop-level detail of ride checks that is useful for developing route restructuring recommendations. In addition, fare-evaders may not be tallied. Amid the COVID-19 pandemic, some transit agencies offered fare-free services, which makes it difficult to obtain ridership data through farebox data.

With the increasing adoption of AFC smart cards that are radio-frequency identification (RFID)-enabled, some transit agencies are able to collect richer data via fareboxes. In metropolitan areas such as New York City and Washington, D.C., smart cards have been widely used with a high penetration rate. For example, Reddy et al. (2009) reported that New York City Transit used farebox data to infer ridership data. AFC-based data collection provided relatively high accuracy and consistency by eliminating the human element in data collection. Reportedly only 3% to 5% of riders in those major cities did not use smart cards and therefore could not be counted (Brakewood, 2014). However, AFC-collected data also needs AVL to gather location information. For example, Lu and Reddy (2012) found that AFC needs to be integrated with AVL data to determine detailed ridership distribution at peak load points.

In addition, a few approaches utilized mobile apps to collect ridership data. For example, DART (2018) utilized the GoPass mobile app for electronic fare payment and data collection. However, the accuracy of such mobile app-collected data is still unclear, as very few studies to date focus on their data quality.

A few studies have developed regression models to estimate ridership based on different contributing factors. For example, Fehr & Peers (2018) combined StreetLight data and on-board passenger survey data to estimate regional ridership data. Similarly, ridership data are estimated and validated with benchmark data collected by fareboxes (Lawson et al., 2021) and APCs (Jung and Casello, 2019). However, such estimation of ridership data was limited as it may be affected by other factors, and there was a lack of analysis on details such as stop-level ridership.

Three studies examined the ridership of vanpool services and demand-response buses. Many vanpool companies submit forms monthly to the overseeing agencies. For example, RTAMS (2020) provided publicly accessible vanpool ridership data for January 2003 through March. However, details on the data collection method and validation procedure were not included in the study. Similarly, other identified studies that include vanpool do not state the data collection methods used (DART, 2018; Toon, 2018).

Ridership Data Collection & Validation Issues

Aging devices pose accuracy issues for ridership data. For example, DART (2018) found aging fareboxes to fail at a regular rate, causing many trips to be completed without the ability to count riders. This leads to undercounting ridership. Similarly, SFMTA (2020) indicated that many buses with older-generation APCs were outdated as new buses with newer-generation APCs began service. The insufficient coverage of new APCs during the transition impeded accurate estimates of crowding.

Given the accuracy issues, researchers have explored diverse approaches to post-process collected data for improving data quality. Saavedra (2010) proposed a quality assurance system and compared a manual survey count in October 2008 by Grand River Transit staff with APC/AVL-collected data. The absolute percentage error was found to be around 11%. Similarly, Chu (2018) developed a tool to save administrative costs of data processing and reporting and to increase the quality of vanpool data on service provided and consumed that was reported to the NTD.

Although about 40 references were identified as relevant to public transit ridership data collection, only 11 studies presented quantitative conclusions on the data accuracy and performance measurements. Few studies applied diverse sampling approaches, estimation methods, evaluation metrics, comparison pairs, and different data sets. For example, Yang et al. (2015) evaluated annual ridership farebox data, while Sound Transit (2015) used daily farebox ridership data. Furthermore, some studies focused on a single route (Tétreault and El-Geneidy, 2010), while others analyzed ridership data at the system level (FCDOT, 2015). In short, studies lack a clear and unified answer on optimal bus ridership data collection approaches.

Summary of Literature Review Findings

Existing studies explored ridership data collection, integration, and evaluation. An interface of the identified literature has been developed to facilitate the exploration of details for those references (see Appendix B). Users can flexibly select and visualize reviewed studies and findings via an interactive user interface. Based on the review of these studies, some key findings by the research team are as follows:

- Farebox technology is the dominant ridership data collection solution deployed by existing transit agencies. Its coverage rate is high in both rural and urban areas.
- APC/AVL technology offers high performance but is relatively expensive to deploy. Its coverage rate is low in rural areas and medium in urban areas. However, some transit agencies (e.g., Hampton Roads Transit) plan to gradually adopt this technology.
- AFC primarily offers high performance and is mainly deployed in metropolitan areas such as New York. AFC data alone does not necessarily capture rider distributions at stops, which often requires integration with an AVL system.
- Manual ride checks are time-consuming but can serve as the benchmark for periodic calibration of other approaches. Manual ride checks are prone to errors introduced by drivers' or ride checkers' attentiveness during manual counting periods.
- A limited number of studies were found regarding vanpool and mobile app-based data collection approaches.

- Aging devices such as fareboxes and APCs can degrade the performance of relevant data collection approaches.
- Unified agreement is lacking on the performance of different ridership data collection approaches due to diverse use of data sampling, estimation, evaluation metrics, and comparison pairs. Few studies were found that evaluate the methodological issues of estimating ridership based on sampled data.

Survey of Transit Agencies in Virginia

Ridership Data Collection Scope and Technical Solutions

All 39 survey responses were accessed and exported from the Qualtrics platform and are listed in Appendix C. Below is a general summary of the responses including incomplete responses to some questions. The approximate number of vehicles in agency fleets ranged from 1 to 312, with an average of 56 vehicles and a median of 31 vehicles. Of responding agencies, 19 served suburban areas, 18 served urban areas, and 16 served rural areas (Note that some agencies may serve more than one type of area).

As shown in Table 3

Table , the number of corresponding respondents was identified for each combination of mode and level of ridership data. These values were divided by the number of all cases (39) to calculate the percentages shown in each cell. The most frequently collected level of ridership data was route level for the bus mode (51.3% of respondents) and for vanpool (7.7%), trip level and route level (tied) for commuter bus (15.4% each) and for trolley-style bus (17.9% each), and trip level for paratransit (33.3%). Respondents frequently did not gather segment-level ridership data. It should be noted that the route-level data reflect unlinked trips, whereas the trip-level data consider multiple linked trips as one trip.

Table 3. Levels of Ridership Data Collection under Different Service Modes (N=39 Agencies)

Service Mode	Level of Ridership Data					
	Stop	Segment	Trip	Route	System	Other
Bus	35.9% (n=14)	5.1% (n=2)	35.9% (n=14)	51.3% (n=20)	30.8% (n=12)	2.6% (n=1)
Commuter bus	12.8% (n=5)	2.6% (n=1)	15.4% (n=6)	15.4% (n=6)	12.8% (n=5)	2.6% (n=1)
Bus rapid transit	0% (n=0)	0% (n=0)	0% (n=0)	0% (n=0)	0% (n=0)	2.6% (n=1)
Trolley-style bus	15.4% (n=6)	0% (n=0)	17.9% (n=7)	17.9% (n=7)	12.8% (n=5)	2.6% (n=1)
Vanpool	0% (n=0)	0% (n=0)	0% (n=0)	7.7% (n=3)	5.1% (n=2)	2.6% (n=1)
Paratransit	12.8% (n=5)	2.6% (n=1)	33.3% (n=13)	15.4% (n=6)	23.1% (n=9)	2.6% (n=1)
Other vehicle types	0% (n=0)	0% (n=0)	0% (n=0)	0% (n=0)	0% (n=0)	2.6% (n=1)

Depending on the technology, data availability can vary (e.g., data may be automatically uploaded from each vehicle to a database once a day, or a technician might enter data from manual paper count sheets once a month). As shown in Table 4, when looking at how quickly new data were available to agency staff, system ridership data and route-level ridership data were most often reported to be available daily (each 43.6% of respondents). Route segment ridership and stop-level boarding/alighting were most often available as needed (23.1% and 30.8%, respectively).

Table 4. Frequency of Data Accessibility (N=39 Agencies)

Frequency of Data Accessibility	Stop-level Boarding/ Alighting	Route Segment Ridership	Route-level Ridership	System Ridership
Daily	23.1% (n=9)	20.5% (n=8)	43.6% (n=17)	43.6% (n=17)
Weekly	10.3% (n=4)	5.1% (n=2)	15.4% (n=6)	20.5% (n=8)
Monthly	12.8% (n=5)	10.3% (n=4)	28.2% (n=11)	28.2% (n=11)
Quarterly	12.8% (n=5)	5.1% (n=2)	15.4% (n=6)	15.4% (n=6)
Annually	12.8% (n=5)	5.1% (n=2)	15.4% (n=6)	17.9% (n=7)
On-demand	30.8% (n=12)	23.1% (n=9)	25.6% (n=10)	25.6% (n=10)

Transit agencies were asked how frequently the different levels of new data were shared outside of their agency (such as with DRPT or with the NTD). As shown in Table 5, none of the transit agencies reported externally sharing data daily. System ridership data was most often shared monthly (59.0% of respondents), while all other levels of data were most often shared on an as-needed basis.

Table 5. Frequency of Data Sharing Outside of Agency (N=39 Agencies)

Frequency of Data Sharing Outside of Agency	Stop-level Boarding/Alighting	Route Segment Ridership	Route-level Ridership	System Ridership
Daily	0% (n=0)	0% (n=0)	0% (n=0)	0% (n=0)
Weekly	0% (n=0)	0% (n=0)	0% (n=0)	2.6% (n=1)
Monthly	2.6% (n=1)	5.1% (n=2)	20.5% (n=8)	59.0% (n=23)
Quarterly	2.6% (n=1)	5.1% (n=2)	7.7% (n=3)	12.8% (n=5)
Annually	2.6% (n=1)	2.6% (n=1)	5.1% (n=2)	28.2% (n=11)
On-demand	35.9% (n=14)	28.2% (n=11)	38.5% (n=15)	17.9% (n=7)

As shown in Figure 1Figure(a), transit agencies were asked what tools their agencies used to collect ridership data. Less than one-third of participants used automated passenger counters (APCs – 30.8%), and less than one-quarter of participants used electronic fareboxes (23.1%) or manual surveys (20.5%). The most frequently selected response was “Others” (41.0%), with respondents most frequently reporting the use of pen and paper, trip sheets, and clickers. Such tools should be considered forms of the “manual survey” option. Other tools mentioned included scheduling software for demand response or paratransit service, on-board mobile data terminals used by drivers, and tickets purchased in advance. Manually reviewing video recordings was the least frequently used method.

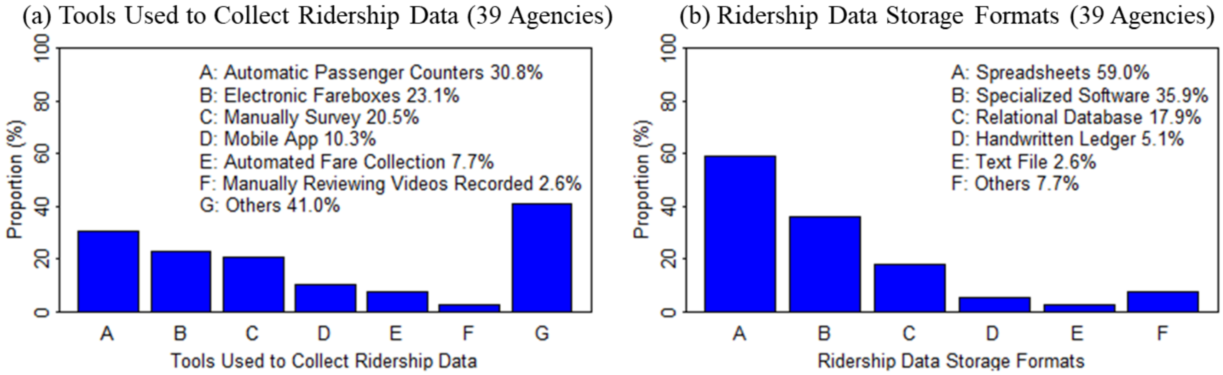


Figure 1. Ridership Data Collection Tools and Storage Formats.

As shown in Figure 1Figure(b), 59% of survey respondents used spreadsheets to store ridership data. Just over one-third used specialized software (35.9%), and 17.9% used relationship databases such as Oracle. Only about 5% used handwritten ledgers, and 2.6% used text files. Other storage formats reported by participants include AgileMile, Database in Routematch, and Electronic Farebox Database.

As shown in Figure 2(a), approximately two-thirds of respondents indicated that ridership data were made publicly available (63.3%). Figure 2(b) shows that 53.3% used specific software tools to analyze ridership data. Those software tools included: Clever, CTS Software, Excel, Routematch, Passio, TransTrack, Hummingbird, Ridecheck Plus APC, TRACI, Jaspersoft, Tableau, Paraplan, Synchromatics, Transitmaster, Avail Technologies, and an agency's own ITS tools.

Transit agencies were asked what data processing steps were applied to validate ridership data. As shown in Figure 2(c), 41% of participants reported comparing ridership with fare revenue or with manual counts. About one-third compared totals across days (35.9%). One-quarter looked for unexplained variations across trips or relied on professional judgment of analysts (25.6% each). Other data processing steps (7.7%) that were listed by the respondents include the following: "*manual sample data is used to calculate PMTs so an expansion process is used on that data," "very accurate count with purchasing tickets in advance and then validating whether passenger took the trip when they check in upon boarding,"* and *"what's given to them from vanpools.*"

Transit agencies were asked how satisfied their organizations were with the quality of ridership data they obtained. As Table 6 reveals, about 46.1% of transit agencies reported being very satisfied or satisfied with APC data, and 54.5% were very satisfied or satisfied with electronic farebox data. Also, 66.6% were satisfied/very satisfied with automated fare collection (AFC) data, and 60.0% were satisfied with mobile app data. Meanwhile, 55.5% of respondents with manual survey data were very satisfied/satisfied.

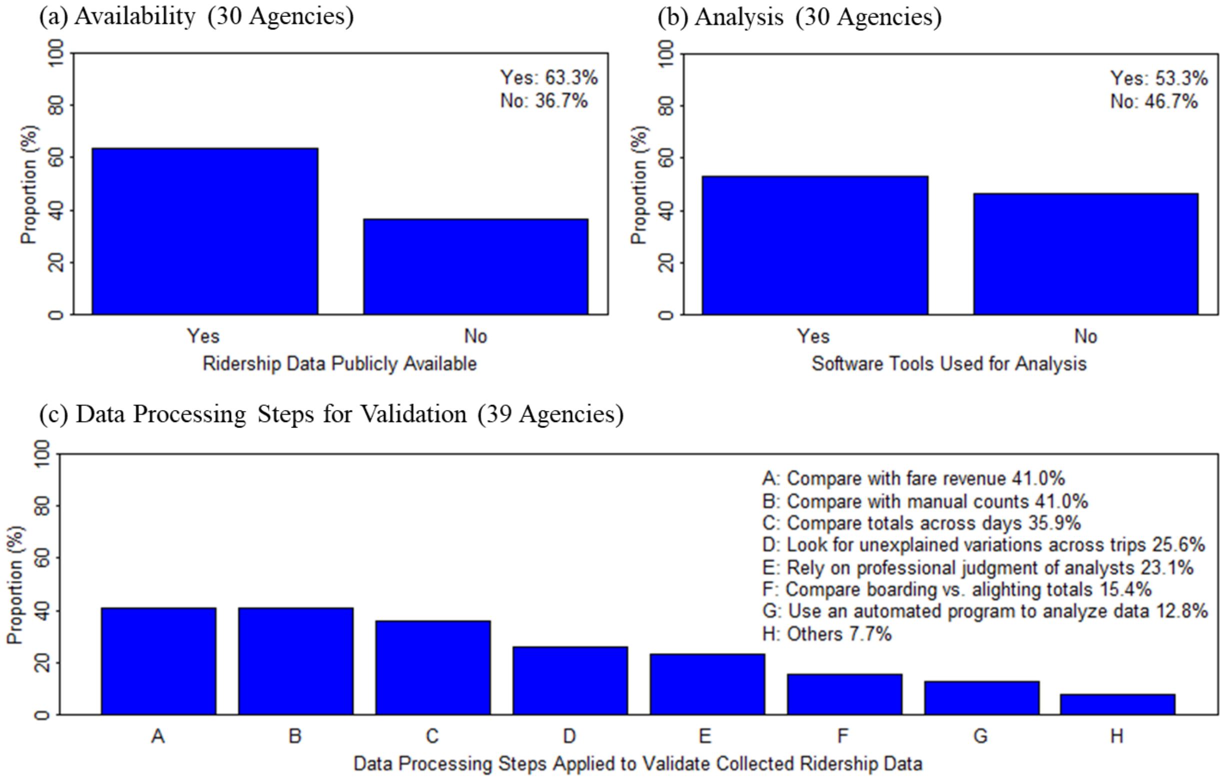


Figure 2. Data Availability, Analysis, and Validation.

Table 6. Satisfaction with Quality of Ridership Data Collected

Type of ridership data (number of responses)	Very satisfied/ satisfied	Neutral	Very unsatisfied/ unsatisfied
Automated Passenger Counter (13 responses)	46.1% (n=6)	30.8% (n=4)	23.1% (n=3)
Electronic Farebox (11 responses)	54.5% (n=6)	18.2% (n=2)	27.3% (n=3)
Automated Fare Collection (3 responses)	66.7% (n=2)	33.3% (n=1)	0.0% (n=0)
Manual Survey (20 responses)	55.0% (n=11)	45.0% (n=9)	0.0% (n=0)
Mobile app (5 responses)	60.0% (n=3)	40.0% (n=2)	0.0% (n=0)

Transit agencies were asked what the primary purposes were for their agencies to collect ridership data. As shown in Figure 3, the most common purposes reported by participants included compiling NTD reports (66.7%), compiling reports for DRPT (66.7%), and identifying their least and most productive routes (61.5%). About 41% reported collecting data to calculate other performance measures, while 35.9% reported using data to identify candidate stops for elimination or addition. Using data to validate travel demand models was reported by 12.8% of participants.

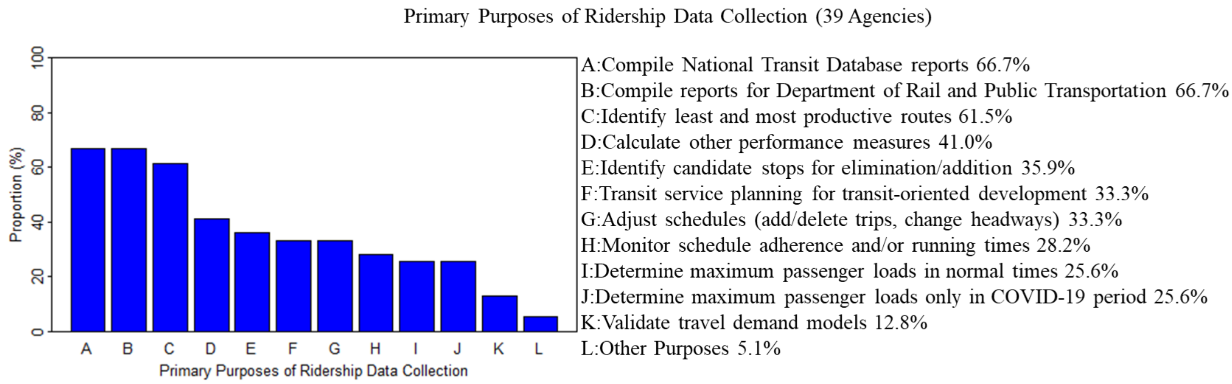


Figure 3. Primary Purposes of Ridership Data Collection.

As shown in Figure 4, more than half of transit agencies reported that raw ridership data were primarily sampled based on all stops/routes (53.3%), 23.3% based ridership data on a sample of routes, and only 10% based ridership data on sampled stops. Other sampling methods mentioned by the respondents included "*based on numbers provided on each van,*" *"Enterprise does this,"* *"for bus unlinked passenger trips (UPTs) based on all routes and PMTs based on sampled trips,"* and *"no sampling method.*"

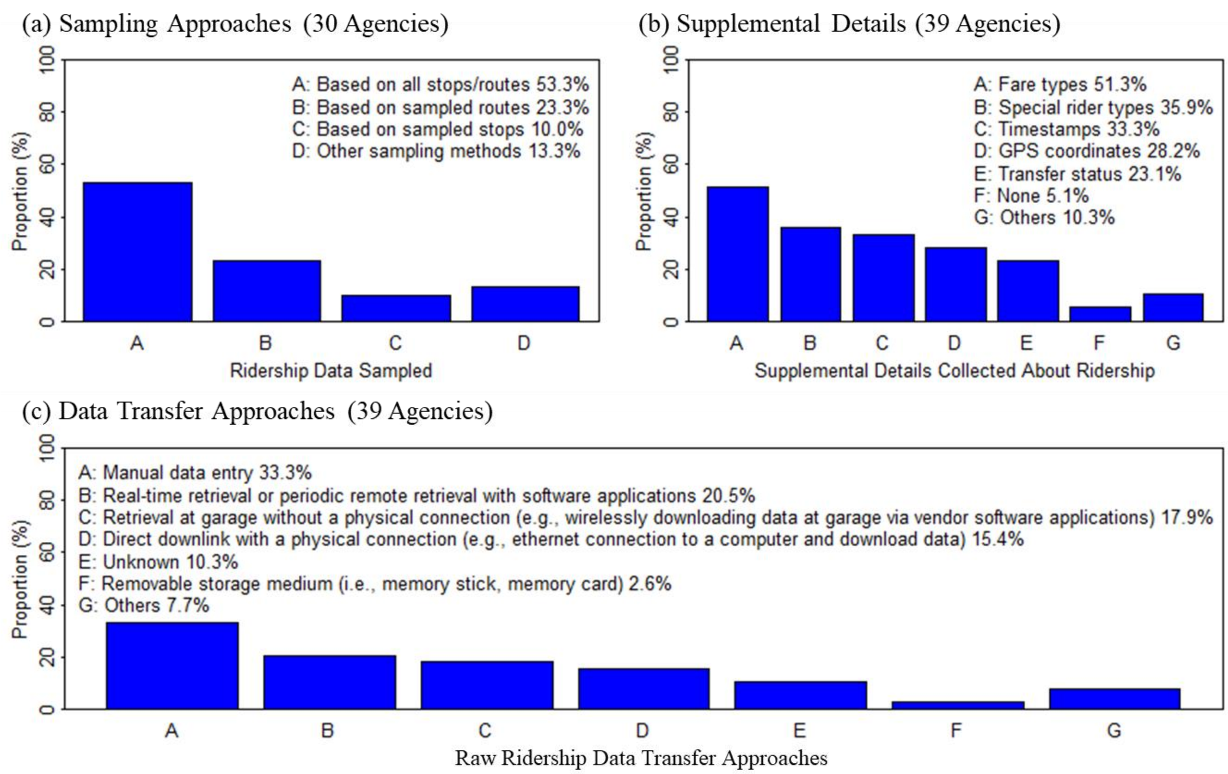


Figure 4. Ridership Data Sampling Approaches, Data Transfer Approaches, and Supplemental Details.

About one-third of the transit agencies indicated that raw ridership data was transferred from data collection devices to storage via manual data entry (33.3%). Also, 20.5% used real-time retrieval or periodic remote retrieval with software applications. Less than 20% used retrieval at the garage with a physical connection (17.9%) or a direct downlink with a physical connection (15.4%). About 10% of respondents did not know how ridership data was transferred to storage, and only 2.6% used a removable storage medium such as a memory stick or card.

Other supplemental details that transit agencies collected about ridership in addition to counts included fare types (51.3%), special rider types (35.9.2%), and timestamps (33.3%). About one-quarter of participants reported collecting location coordinates (28.2%) or transfer status (23.1%). About 5% reported collecting no other details about ridership. Other details collected and mentioned by participants included: purpose, seasonal usage, van size, open seats, and route.

Automated Passenger Counters

As shown in Figure 5(a), 14 transit agencies reported that they had APCs on at least some vehicles used in major service types. For 7.1% of respondents, only 1-25% of their fleet for major service types was equipped with APCs. Two categories each had 14.3% of transit agencies: those with APCs on 26-50% of their fleet and those with APC on 51-75% of the fleet. The majority of agencies (64.3%) with APC technology reported 76-100% of their fleet was equipped with APCs.

One agency reported none of its fleet used in major service types had APCs. However, it actually deployed APCs, based subsequent answers. Thus, the sample size was 15 for Figure 5(b) and 5(c). Fifty-three percent of transit agencies reported all vehicles were equipped with APCs, while 6.7% reported equipped vehicles were rotated between routes or equipped vehicles were dedicated to selected routes. Similarly, 60% of transit agencies indicated the technology was based on infrared light. Twenty percent reported that the technology was based on video, and 6.7% reported it was based on Bluetooth/Wi-Fi. Other technologies mentioned included Hella APC (a video-based device), 3D video, and tablets.

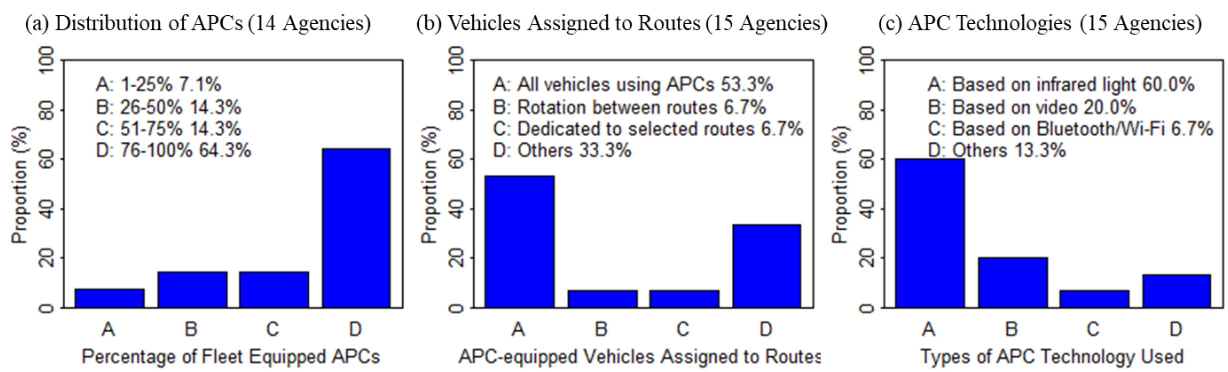


Figure 5. Fleet Equipped with APCs, Route Assignment for Equipped Vehicles, and Types of APC Technology.

Electronic Fareboxes

As shown in Figure 6(a), 10 transit agencies reported that they used electronic fareboxes and indicated that 76-100% of their fleet were equipped with electronic fareboxes. For the route assignment question, one agency selected the "Other" option but did not provide any further information; thus, the agency size is 11 for Figure 6(b).

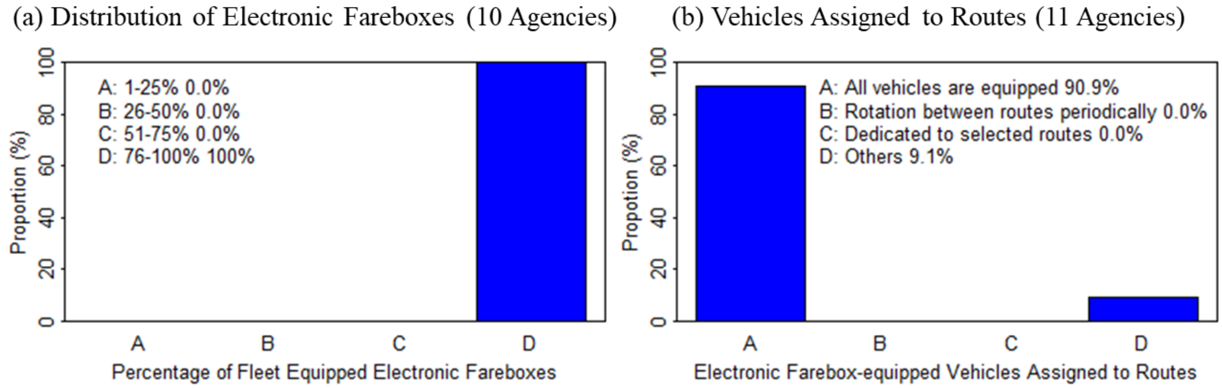


Figure 6. Fleet Equipped with Electronic Fareboxes and Route Assignment for Equipped Vehicles.

Automated Fare Collection Devices

As shown in Figure 7, 50% of the four agencies with AFCs indicated that 76-100% of their fleets were equipped with AFCs. One respondent indicated that 1-25% of their fleets were equipped with AFCs, and one respondent indicated 51-75% of their fleets were equipped with AFCs. A fifth respondent selected the "Other" option for the route assignment question, making the sample size 5 in Figure 7(b), while it is 4 in Figure 7(a). The rest indicated that all vehicles were equipped with AFC devices, but this should be interpreted with caution, as it is not consistent with the answers shown in Figure 7(a).

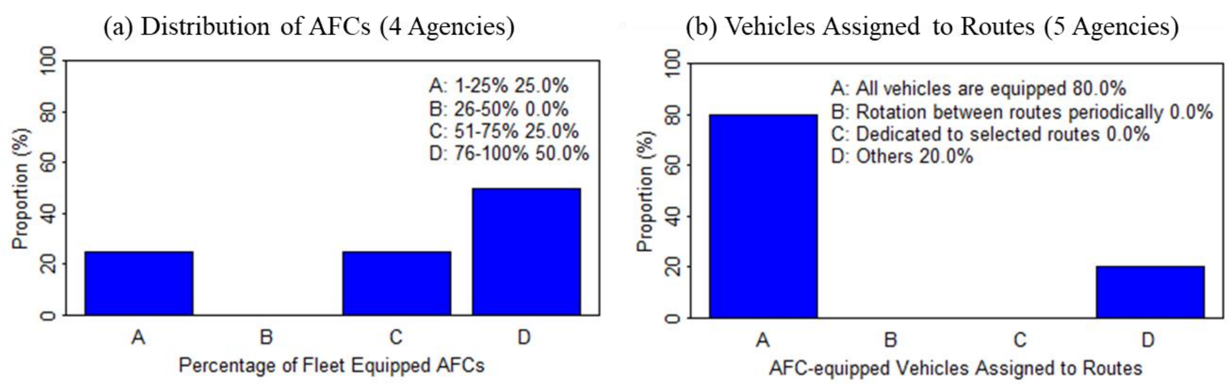


Figure 7. Major Service Types Equipped with AFCs.

Manual Surveys

As shown in Figure 8, about 7% of agencies reported that their organization never conducted manual surveys to collect ridership data. About 10% conducted such surveys weekly, and 3.4% conducted them quarterly. 27.6% conducted manual surveys annually, while 51.7% conducted them at other timeframes, with the most common response being as needed/required or when requested. About 74.1% of agencies indicated that these surveys were conducted onboard vehicles, and 37.9% conducted these surveys on the majority of their routes. This was followed by 27.6% of agencies who only conducted manual surveys on 1-25% of their routes.

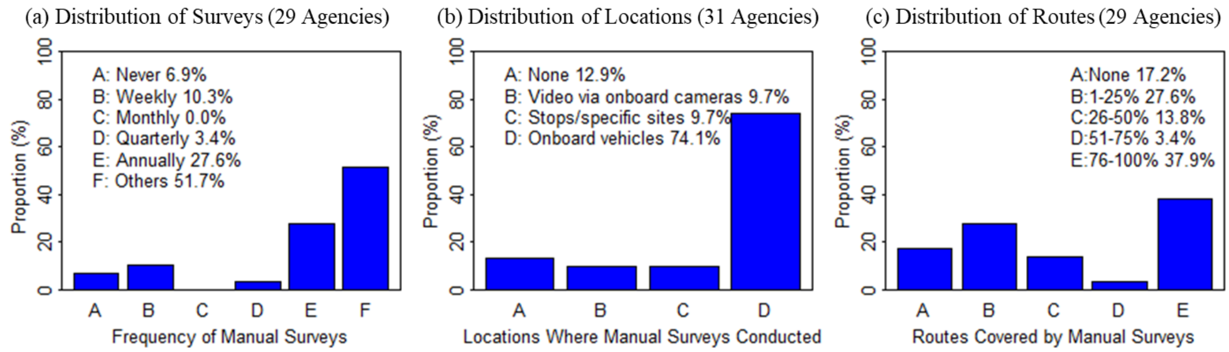


Figure 8. Details of Manual Surveys.

Sampling & Estimation Techniques

For each raw data collection method, the number of transit agencies (out of all 39 responses) that indicated they used sampling techniques is shown in Table 7. There were 11 responses for APC passenger counts, 10 for farebox passenger counts, 22 for manual passenger counts, 5 for mobile ticketing passenger counts, 3 for AFC revenue data, 18 for mobile app revenue data, and 13 for order data. Transit agencies were most likely to indicate that they used sampling techniques on order data (61.5%), manual passenger counts (54.5%), or farebox revenue data (50.0%).

Table 7. Sampling Techniques of Raw Data Collection

Type of Ridership Data	Yes	No
APC passenger counts	45.5% (n=5)	54.5% (n=6)
Electronic farebox passenger counts	30.0% (n=3)	70.0% (n=7)
Manual passenger counts	54.5% (n=12)	45.5% (n=10)
Mobile ticketing passenger counts	20.0% (n=1)	80.0% (n=4)
AFC revenue data	0.0% (n=0)	100.0% (n=3)
Farebox revenue data	50.0% (n=9)	50.0% (n=9)
Mobile app's revenue data	0.0% (n=0)	100.0% (n=4)
Order data (e.g., reservation records of paratransit)	61.5% (n=8)	38.5% (n=5)
Other (responses included "online ticket purchase," "verification upon boarding," and "unsure if there is a sampling technique")	33.3% (n=1)	66.7% (n=2)

With regard to how agencies obtained long-term ridership estimates, regardless of the data type, many participants chose the "other methods" response. Within that "other methods" category, however, the dominant specific answers were "none," "we do not estimate future ridership," "we count all passengers," and "actual data collected daily." Responses of that nature were excluded from the "Other methods" column in Table 8, since those agencies did not create estimates of long-term ridership. The remaining two "Other methods" responses specified "trend analysis." As shown in Table 8, regression methods seemed to be used the least, regardless of data type, and scaling methods used the most.

Table 8. Long-term Ridership Data Estimation

Type of Ridership Data	Scaling method	Regression method	Weighted average	Other Methods [a]
APC passenger counts	50.0% (n=4)	12.5% (n=1)	37.5% (n=3)	0.0% (n=0)
Electronic farebox passenger counts	60.0% (n=6)	20.0% (n=2)	20.0% (n=2)	0.0% (n=0)
Manual passenger counts	57.1% (n=8)	14.3% (n=2)	21.4% (n=3)	7.1% (n=1)
Mobile ticketing passenger counts	75.0% (n=3)	25.0% (n=1)	0.0% (n=0)	0.0% (n=0)
AFC revenue data	60.0% (n=3)	0.0% (n=0)	40.0% (n=2)	0.0% (n=0)
Farebox revenue data	70.0% (n=7)	20.0% (n=2)	10.0% (n=1)	0.0% (n=0)
Mobile app's revenue data	50.0% (n=1)	0.0% (n=0)	50.0% (n=1)	0.0% (n=0)
Order data (e.g., reservation records of paratransit)	66.7% (n=6)	0.0% (n=0)	22.2% (n=2)	11.1% (n=1)
Other data	60.0% (n=3)	20.0% (n=1)	20.0% (n=1)	0.0% (n=0)

[a] *Note*: *See Appendix D for a summary of these other responses.*

Table 9 shows most participants (85.7%) responded "not applicable" when asked how they integrated data for longer-term ridership estimation if using multiple data sources for their primary service.

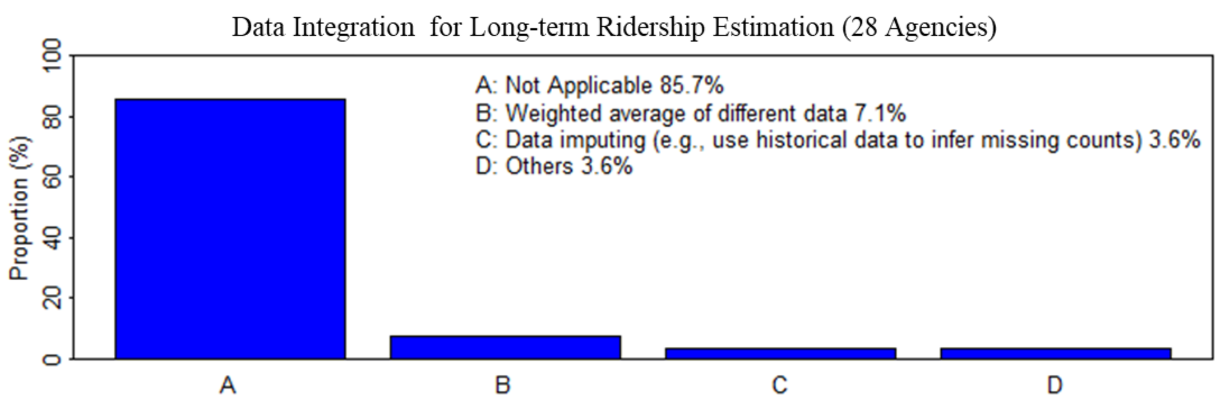


Figure 9. Data Integration for Long-term Ridership Estimation.

National Transit Database Reporting and Tracking

As shown in Figure 10, 13 respondents indicated that their agencies reported full data to NTD (46.4%), with the same number indicating their agencies reported reduced data, while two respondents' agencies did not have NTD reporting duties. Participants were most likely to report tracking UPT for the NTD (89.3%), followed by total PMT (50.0%). As shown in Figure 10(c), most respondents indicated that they counted all passengers without sampling (67.8%) for the NTD, and 17.9% used the NTD-provided sampling algorithms. About 10% used a non-NTD sampling procedure with a qualified statistician. For respondents' descriptions of the steps their agencies took to validate ridership data for NTD reporting purposes, please see Appendix E.

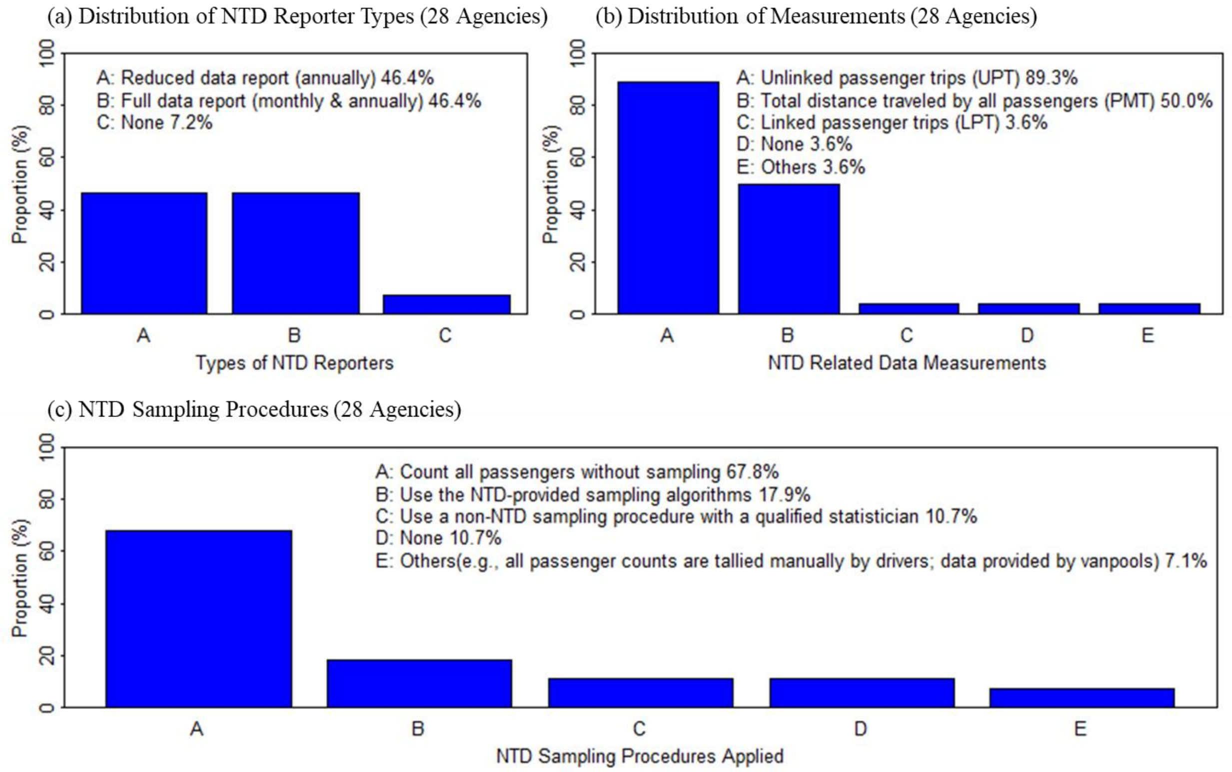


Figure 10. NTD Reporting and Tracking.

When asked what steps were taken to validate ridership data for NTD reporting purposes, respondents' answers fell into the following major categories (ranked by response frequency):

1) Validation by comparison with historical data (e.g., previous month/previous year data),
2) Reviewed by internal staff,
3) Validation by comparing performance metrics (e.g., reasonableness checks such as passengers per hour),
4) Validation by comparison with other sources of data (e.g., manual counts vs. APCs); and
5) Checked by external agencies (e.g., a consultant firm).

Changes to Ridership Data Collection Processes

As shown in Figure 11, most participants (82.8% of 29 agencies) indicated that their agencies had not made any changes to their ridership data collection processes in recent years. About 7% had expanded electronic data collection efforts, and 3.4% had expanded data collection scale or improved their ridership data estimation approaches. The one agency under "Others" noted the data collection process for its light rail switched from manual sampling to full APC counts. The majority of participants indicated that prior to the COVID-19 pandemic, their agencies had not planned to change their ridership data collection process by 2022 (82.1%).

About 7% indicated that they had planned to expand electronic data collection efforts, and about 10% had other changes planned (e.g., “currently working to get our bus APC certified by FTA for NTD reporting,” “Purchase/install APC on all buses,” and “We will be moving to Clever devices from our current APC”).

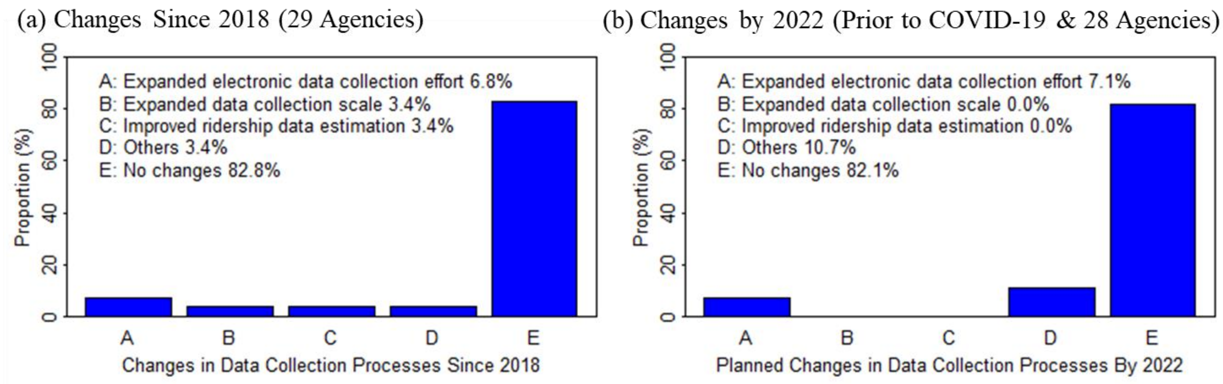


Figure 11. Changes to Ridership Data Collection Processes.

Only one survey respondent indicated that their agency planned to change its ridership data collection process in response to the impact of the COVID-19 pandemic. The respondent stated that the FTA required the agency to shut down its manual sampling activities for PMT from March to July 2020.

Evaluation of Transit Ridership Data Collection Approaches

Based on the survey results, the research team further contacted a subset of transit agencies to acquire historical ridership data for analysis of the potential estimation error if a sampling approach was used instead of reporting the raw annual ridership. It should be noted that all five agencies that provided data collect daily ridership using a specific approach (e.g., manual count, farebox data, and/or APC data). These collected data were reviewed by the agencies and used for reporting. Since the research team obtained full annual data as the benchmark, it allowed us to test the possible estimation error as if only a subset of data was collected as a sample. The following sections show the analysis results based on data from different transit agencies in Virginia. The actual transit agency names were coded as Agency A to Agency E.

Comparisons between APC and Farebox Data

Both FY18-19 APC and farebox data from the fleet of Transit Agency A were available. This agency operates a fleet of about 50 vehicles that cover both regular and non-regular bus routes. Many non-regular routes only operated during a specific time, e.g., summer months. The access to these two types of data from regular bus routes facilitated a direct comparison between them. In addition, the provided data were recorded daily for different routes. Thus, we compared the farebox data and APC data for each route, and the following equation was used to show the relative difference between farebox data and APC data for a set of selected routes.

$$Relative\,Difference = \frac{Daily\ APC\ \text{-}\ Daily\ Farebox}{Daily\ Farebox} \times 100\% \qquad (4)$$

We selected 20 major routes operated by the agency and computed the relative difference between the APC data and farebox data. Figure 12 shows the comparative results. Among the compared routes, we can see that the daily APC data from some routes–i.e., Routes 1 to 6–were systematically higher than the corresponding farebox data. The APC data were about 10% to 30% more than farebox data for Routes 1, 2, 3, and 5. The discrepancy was over 40% for many records from Routes 4 and 6. However, the APC data tended to be lower than the farebox data for some other routes, such as Routes 18 and 20, for which APC data were 5% to 10% lower than most farebox records. In addition, for some routes, the APC data and farebox data did not exhibit clear differences in some cases (e.g., Routes 11 and 12).

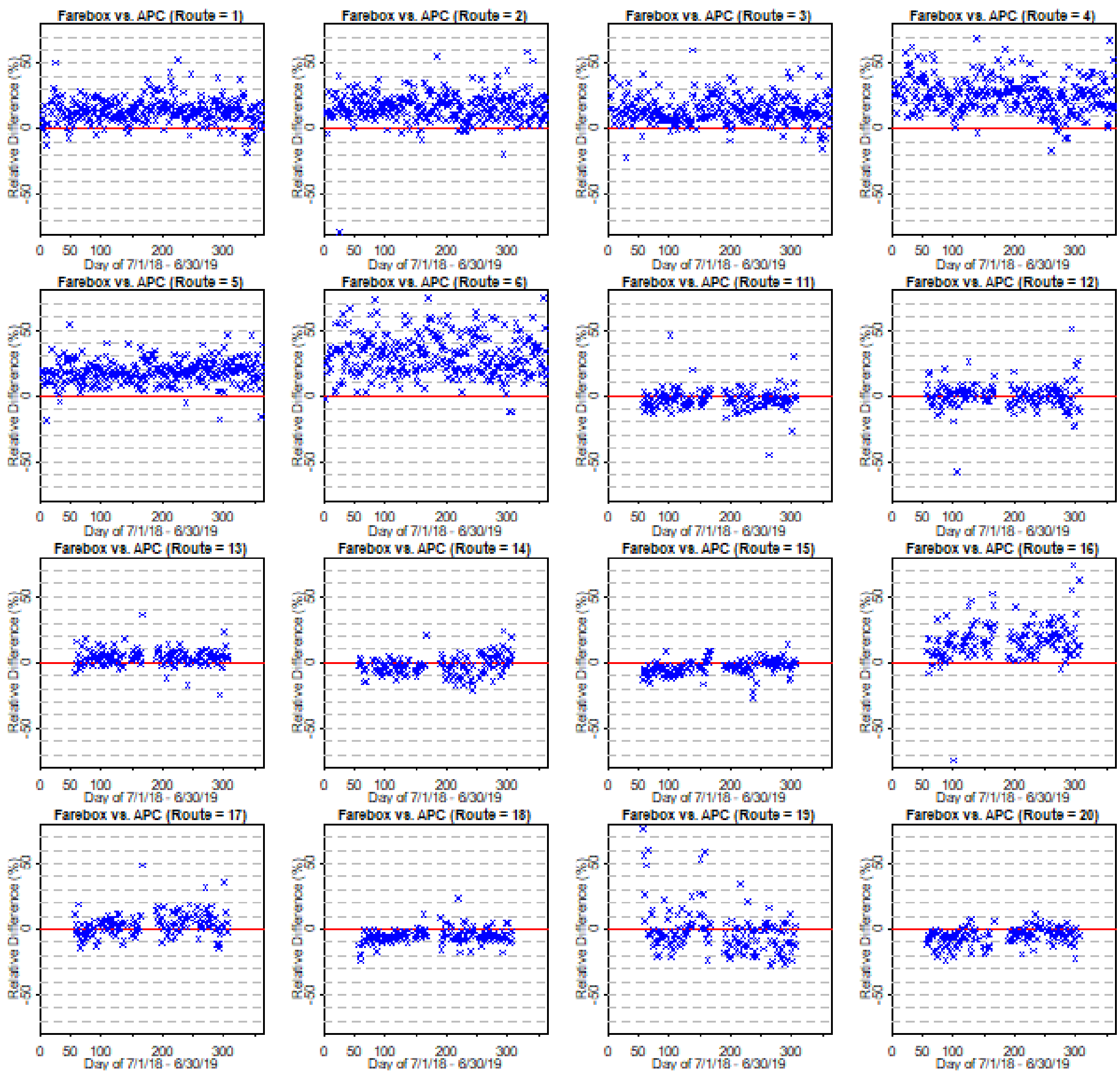


Figure 12. Relative Difference between Daily APC and Farebox Data of Different Routes (Note: X's above the Zero Line Indicate that APC Data is Higher Than Farebox Data).

Since farebox-based ridership data were mainly derived based on the collected fares, accurate statistics on the collected fares will help derive reliable counts. Nevertheless, fare discounts for some riders may reduce the accuracy of farebox-based ridership estimation. Carefully verifying and updating the rider counts associated with those using coupons can help reduce the error. In contrast, even if the APCs were well-calibrated and certified, the APC counts may be subject to errors due to several factors. For example, in an interview with a project manager at Agency A, it was mentioned that the APC data can be inflated by bus drivers' boarding and alighting (such as for breaks), and that different drivers may board and alight different numbers of times. Also, passengers momentarily boarding a bus to make inquiries (such as asking the driver which bus to take) may also be counted by APCs, even though they may not actually ride that bus.

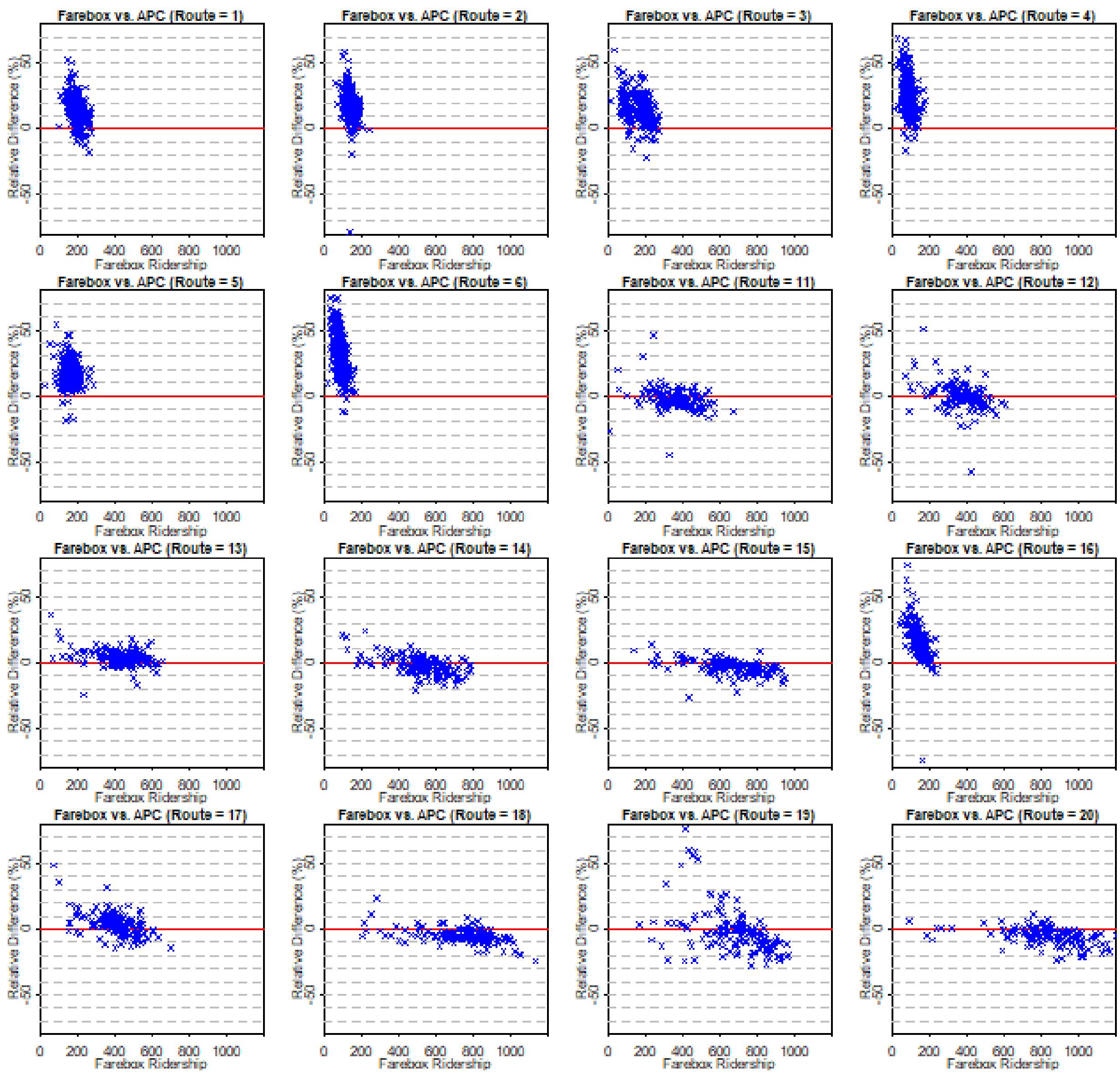


Figure 13. Relative Difference vs. Farebox-based Ridership of Agency A.

Following the exploration of the day-to-day discrepancies between APC data and farebox data, we explored how the discrepancies vary with respect to route-level ridership. Figure 13 shows the results. For some low-demand routes (e.g., Routes 1 to 6), we can see that the relative difference tends to be larger for the days when farebox ridership is smaller. This indicates that proportionally APC data will be notably higher than farebox data. For the other routes, there was no clear pattern between the level of farebox ridership and relative discrepancy.

Ridership Estimation Error based on Different Sample Sizes

Estimation Error Based on Data from Agency B

Agency B provides services in southwestern Virginia and has approximately 50 vehicles in its fleet. The system typically operates during the weekdays. Figure 14(a) shows the collected daily ridership of the system in the past three fiscal years. These trip count data were collected via manual counting by the route drivers, kept on written tally sheets, and returned with daily paperwork for data entry. The aggregated annual ridership was used for reporting. As the full data are available, this provides the benchmark for comparing the estimated ridership based on only a subset of the manual count with the actual annual ridership.

Based on the estimation method introduced in the Methods section of this report, Figure 14(b) shows the estimation errors for FY17-18 and FY18-19. Specifically, to estimate the FY17-18 ridership, FY16-17 data were used to determine the minimum sample size, which was found to be 11 weeks, based on the procedure described in the Methods section. Likewise, to estimate the FY18-19 ridership, FY17-18 data were used to determine the minimum sample size, which was determined to be five weeks. We can see that due to the variation in ridership during the previous year, the minimum required sample size for estimating a target year's ridership can change. Using 10 random sampling experiments and equation (3), estimation errors were calculated for each estimation experiment. Figure 14(b) shows that the error was between -4% and 7% for FY17-18 and between -15% and 4% for FY18-19.

Figure 14(c) shows how the errors would change among 10 sampling experiments if FY17-18 ridership were estimated based on different sample sizes. As the minimum sample size needed was 11 weeks of data, reducing the sample size to five weeks of data tended to raise the estimation error, with errors in some experiments reaching 10%. In contrast, most errors were found to be less than 5% after increasing the sample size to 15 weeks of data.

To explore possible errors due to the random sampling, the estimation experiments were repeated 20 times under each assumed sample size (i.e., five weeks, 10 weeks, and 15 weeks), and the results are shown in Figure 14(d). The boxplot shows how the 20 calculated values for percentage error γ_k (k =1, 2, …, 20) can change. The median of these percentage differences is shown as the thick line in the box. The top and the bottom of each box show the 25th (Q1) and 75th (Q3) percentiles of these percentage errors, respectively. Their difference represents the interquartile range (IQR). The dashed lines indicate Q1-1.5*IQR and Q3+1.5*IQR, respectively. Any values beyond the dashed lines are considered to be outliers / extreme cases (shown in circles in the chart). If more sampled data were used in estimation, the estimation error was reduced, as more error values are centered around zero. There is a higher chance of obtaining a large variation among repeated experiments if a smaller number of weeks (e.g., W =5) was sampled for estimating ridership. Because of periodic variability in daily ridership, small samples are likely to be tied to larger variations in ridership.

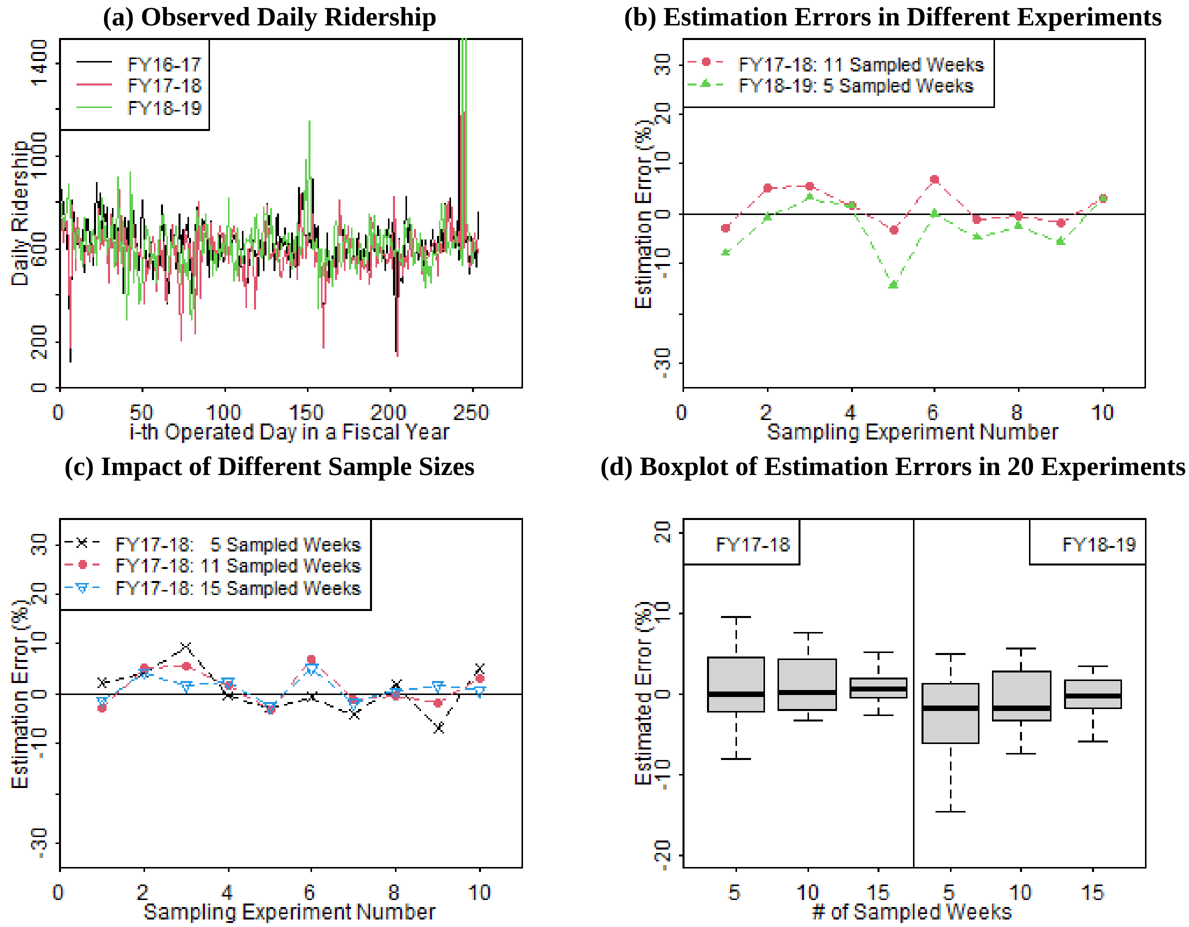


Figure 14. Analysis of Ridership from Agency B (Ridership Collection: Manual Count; Boxplot: Error Distributions of 20 Repeated Experiments).

Estimation Error Based on Data from Agency C

Agency C has approximately 12 vehicles in its fleet, and its weekday ridership data were obtained for analysis. These ridership data were gathered daily by drivers using manual clickers. As shown in Figure 15(a), the ridership levels for FY16-17 and FY17-18 were similar until near the end of FY17-18, whereas the daily ridership in FY18-19 was about one-third of that for most of the previous two years. There may have been a substantial service change amid the last period of FY17-18, and the subsequent FY18-19 daily ridership continued at the lower level. We used the FY16-17 and FY17-18 data as the basis to determine the sample sizes for estimating ridership in FY17-18 and FY18-19, respectively. Note that this may not be the best option due to significant changes between FY17-18 and FY18-19. As no data were available for the period following the changes, we did not separate pre- and post-change data for estimating ridership. The estimation errors of 10 experiments using those sample sizes are shown in Figure 15(b). We can see the errors of FY17-18 estimation in some experiments were over 30%. With six weeks of data used for FY18-19 estimation, most of the errors were within ±10%. For the same year, Figure 15(c) clearly shows that increased sample sizes will help reduce the estimation errors. Increasing sample size from 10 to 15 weeks did not notably change the estimation errors. As with the results based on Agency B's data, with repeated sampling experiments, Figure 15(d) shows the distributions of estimation errors for a given sample size. The results also suggest that increased sample sizes help reduce estimation errors and their variances.

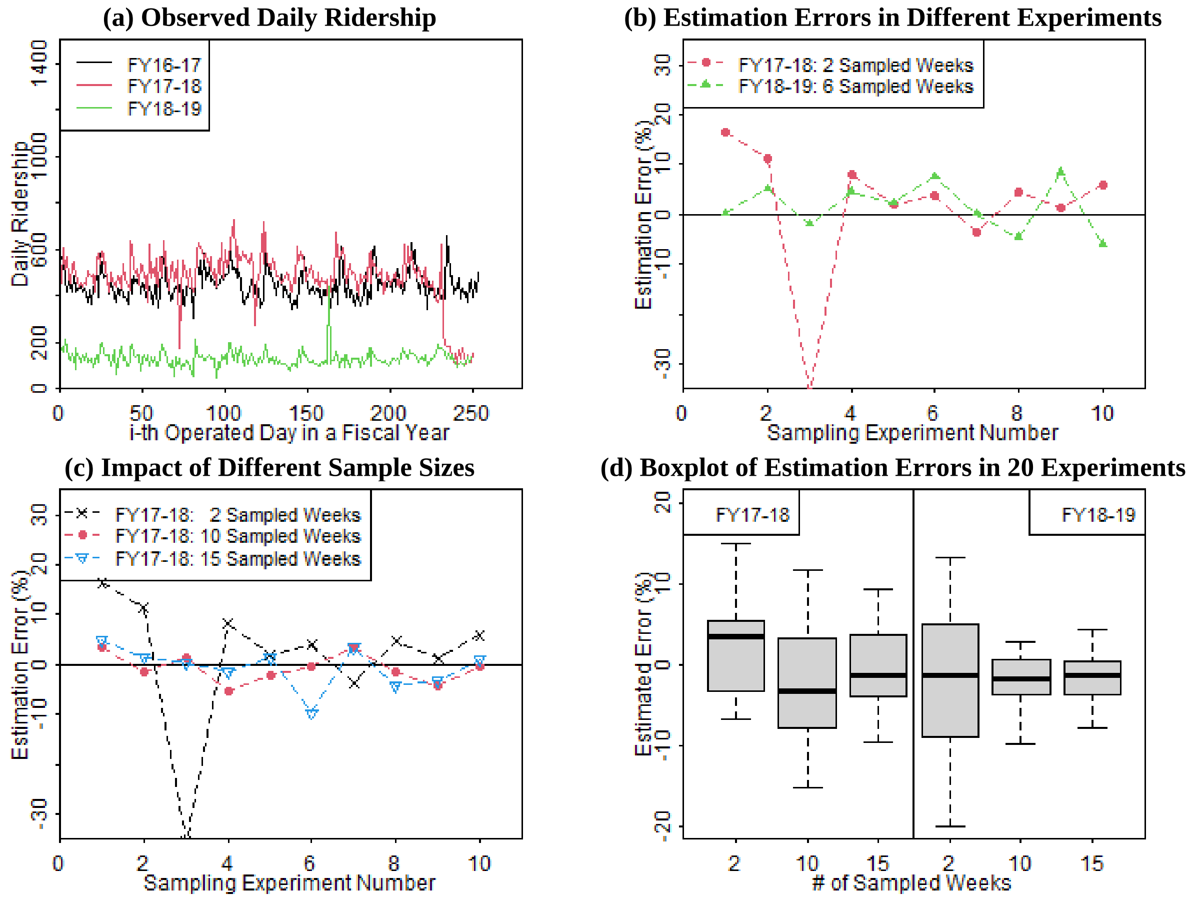


Figure 15. Analysis of Ridership from Agency C (Ridership Collection: Clicker; Boxplot: Error Distributions of 20 Repeated Experiments).

Estimation Error Based on Data from Agency D

Agency D has about 150 vehicles in its fleet, and the daily ridership totals collected on weekdays are shown in Figure 16(c). Its service covers both urban and suburban areas in Virginia. The agency uses farebox data for reporting ridership data, and its daily ridership of over 9,000 is significantly higher than the daily ridership of Agency B or C. We applied the same estimation approaches to examine the estimation errors should sampled daily farebox data be used for estimating annual ridership. Based on the prior year's data, we determined the sample size for a target year's estimation. As shown in Figure 16(b), we determined that four weeks of data were needed for estimating both FY17-18 ridership and FY18-19 ridership. In the repeated sampling experiments, it was found that the estimation error fluctuated between -8% and 8% for FY17-18 and between -11% and 8% for FY18-19. When the sample size was increased to 10 and 15 weeks, Figure 16(c) shows that the estimation errors were reduced, and Figure 16(d) shows the reduction in variation of the estimation errors. Consistent with the manual counting scenarios, when 10 or more weeks of data were sampled for estimation, the errors tend to be within ±10% as confirmed by repeated sampling experiments.

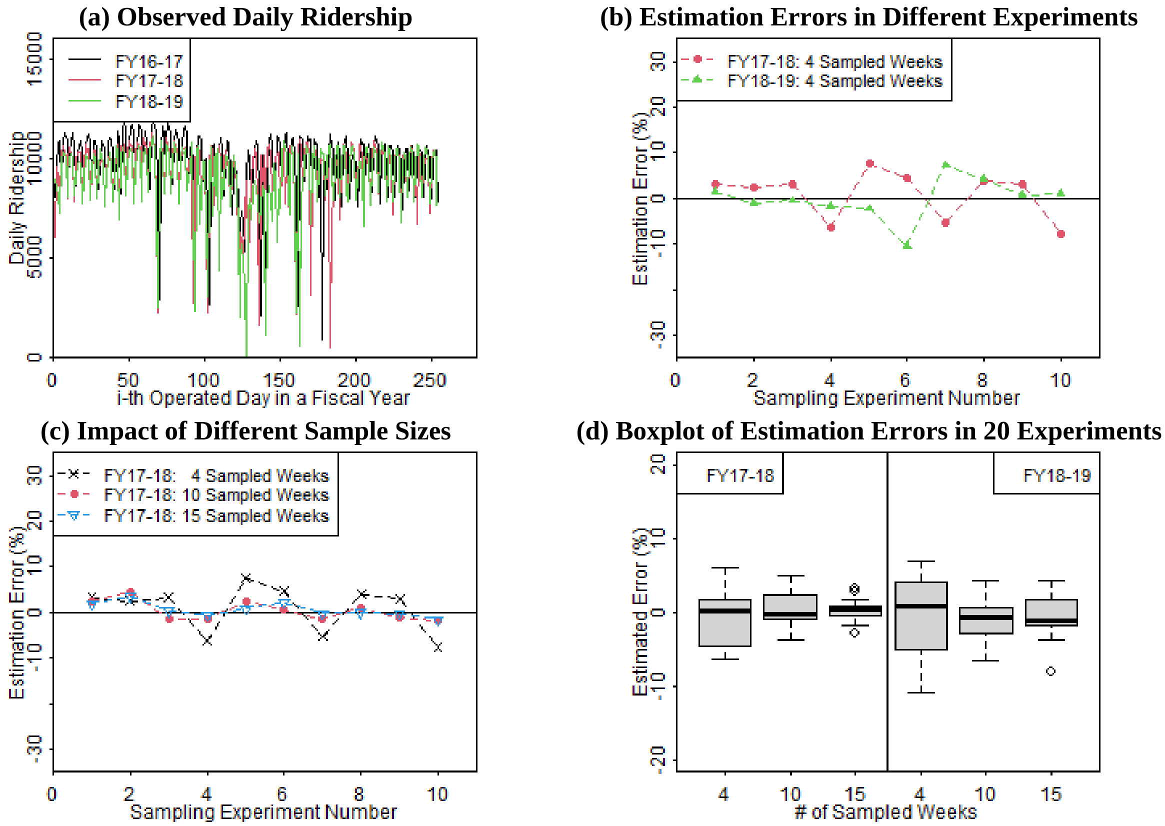


Figure 16. Analysis of Ridership from Agency D (Ridership Collection: Farebox; Boxplot: Error Distributions of 20 Repeated Experiments).

Estimation Error Based on Data from Agency E

Transit Agency E is one of the largest transit service providers in Virginia and has over 280 vehicles in its fleet providing fixed-route bus service and demand-response paratransit service, among others, on both weekdays and weekends. For fixed-route bus service, the agency relies on farebox data for ridership data collection. For demand-response service, its ridership data are collected through scheduling software. We obtained three years of daily ridership data for the demand-response service and for 10 representative fixed routes. Although the services are usually available on weekdays and weekends, the levels of ridership typically differ between weekdays and weekends for each type of service. For example, Figure 17(a) illustrates the ridership by fiscal year for the demand-response service. The weekly cyclic pattern is clear: weekdays often maintain higher demand, whereas weekend demand is much lower. Taking FY18-19 as an example, the average weekday ridership and its standard deviation are 1,096 and 172, respectively. In contrast, the average weekend ridership and its standard deviation are 418 and 110, respectively. The ridership for the fixed-route bus service also shows similar weekly cyclic patterns (Figure 17[b]).

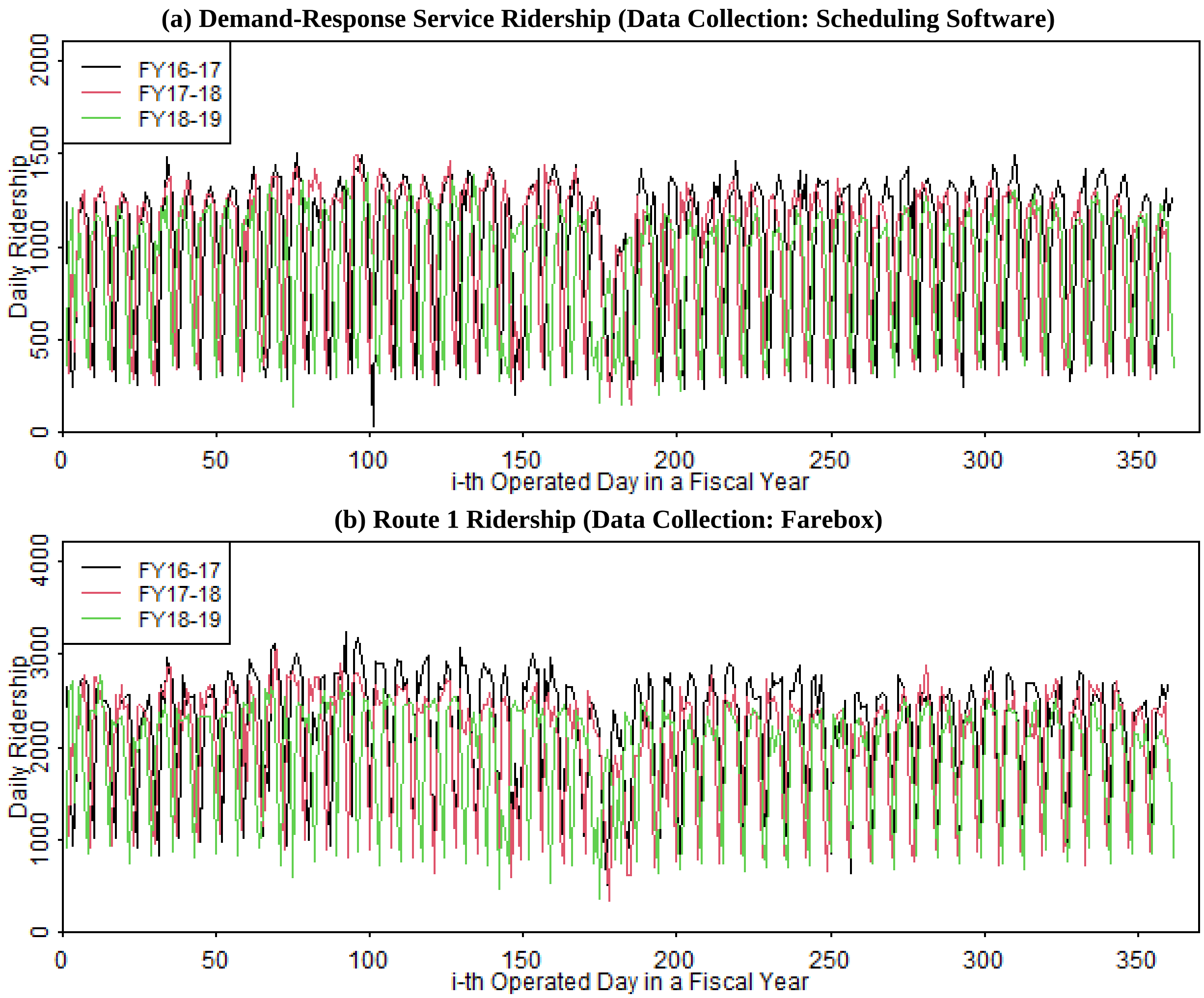


Figure 17. 3-Year Ridership of the Demand-Response Service and a Selected Fixed Route of Agency E.

Due to the notable difference between weekday and weekend ridership, it is rational to separately consider them in estimating annual ridership based on the introduced sampling approach. In the subsequent analysis, we use only the weekday data for illustrating the possible estimation errors. As performed in analyses for other agencies, the annual weekday ridership was estimated for the demand-response service of Agency E. The actual ridership and the estimate based on sampled data were compared to determine the estimation error. Based on the weekday data in Figure 18(a), we determined that at least three weeks of sample data would be needed for each target year (FY17-18 and FY18-19). With three weeks of randomly sampled data, the estimation errors of 10 repeated experiments (Figure 18(b)) were found to be between -5% and 10% for FY17-18 and between -6% and 6% for FY18-19. We further examined the effect of sample size and the distributions of estimation errors under different sample sizes. The results are shown in Figure 18(c) and Figure 18(d), respectively. Consistent with previous findings, increasing sample size helps reduce the estimation error and its variation, although the latter was less apparent for FY18-19 data. Most of the estimation errors are within ±10% when 10 or more weeks of data were sampled for estimating annual ridership.

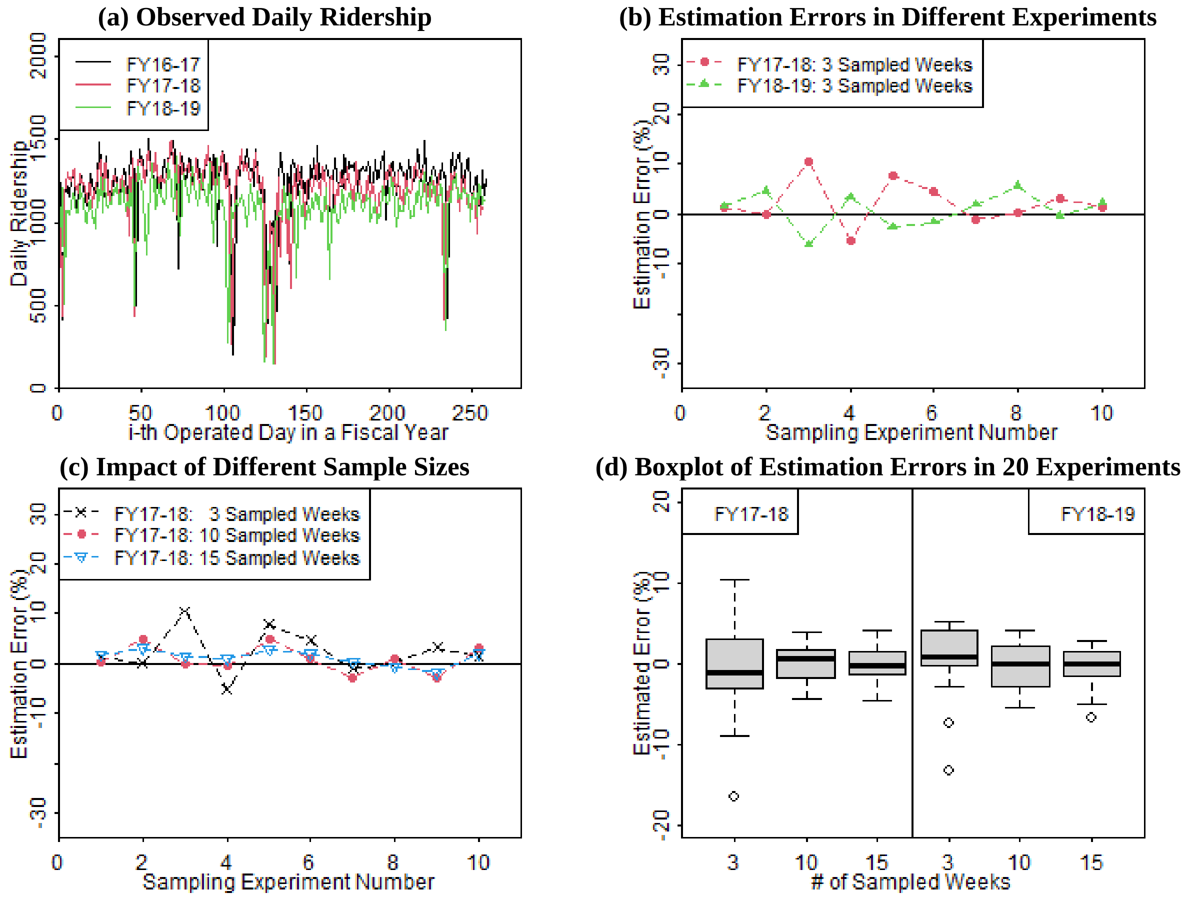


Figure 18. Analysis of Weekday Demand-Response Service Ridership from Agency E (Ridership Collection: Scheduling Software; Boxplot: Error Distributions of 20 Repeated Experiments).

Because ridership data for 10 individual routes were available, we conducted the analysis for each route. The sampling approach was applied to each route, and the estimated annual ridership (based on all weekdays) was compared with the actual observation. An exploratory analysis found that these 10 routes showed different levels of demand. Thus, we selected three routes representing low-, medium-, and high-demand scenarios for testing the estimation errors. The actual (farebox) weekday ridership data of the selected routes are shown in Figure 19(a), (c), and (e). The corresponding estimation errors based on 2-week, 10-week, and 15-week sample data are shown in Figure 19(b), (d), and (f). Despite the differences in demand levels, the estimation error for each route is reduced if an increased sample size is used. As shown by these boxplots, if sample data of 10 or more weeks were used, most of the estimation errors are within ±5%. When the ridership data from all 10 routes is aggregated as "system ridership" (Figure 19(g)), similar estimation error distributions are obtained (Figure 19(h)). If only 2 weeks of sample data were used for estimation, there would be a risk of obtaining errors beyond ±10% (i.e., the circles in Figure 19(h)). The analysis results based on data from these agencies suggest that the number of weeks of ridership data used as a sample is critical for estimating annual ridership, regardless of the scale and demand levels of the system.

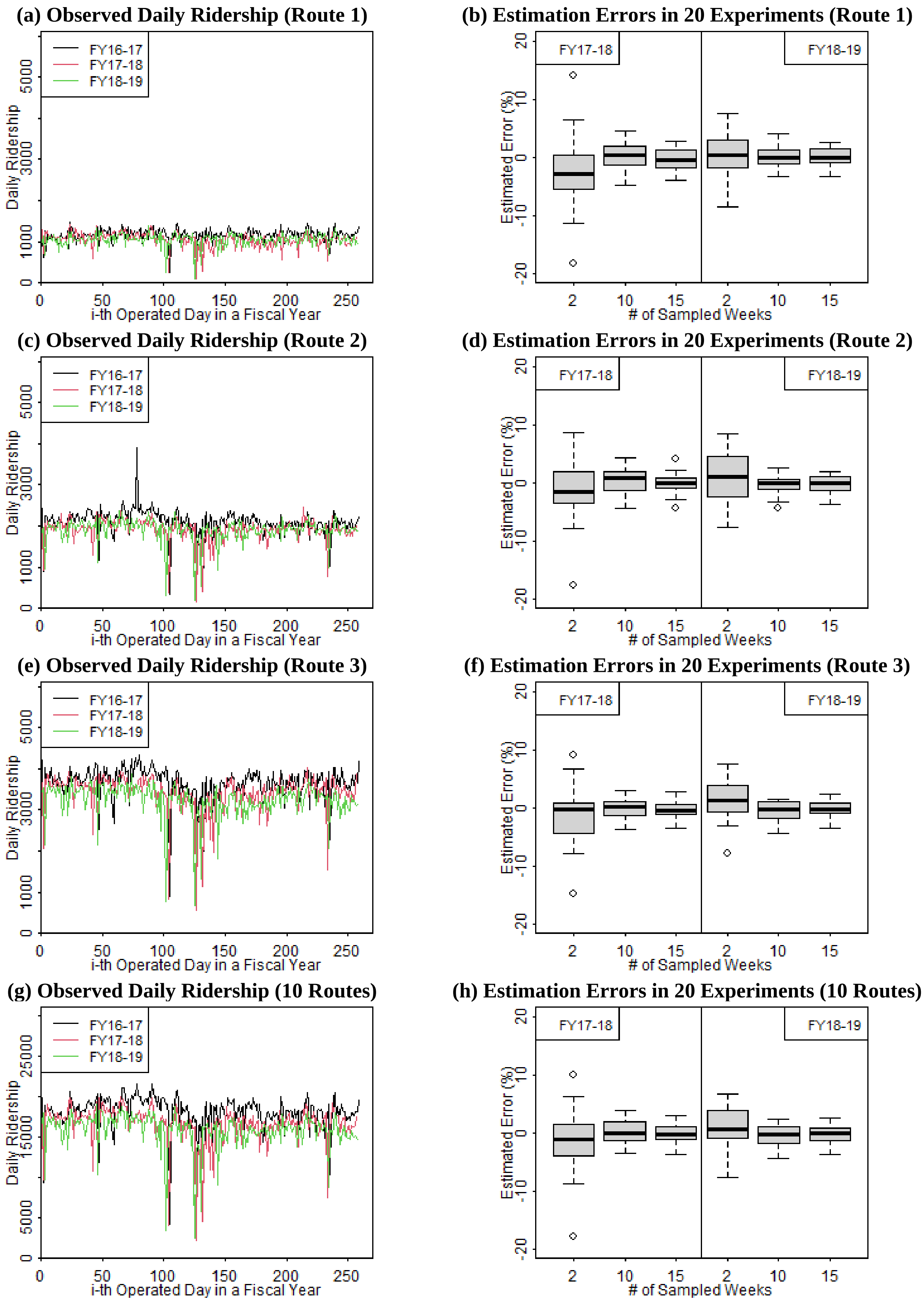


Figure 19. Analysis of Ridership of Selected Routes from Agency D (Ridership Collection: Farebox; Boxplot: Error Distributions of 20 Repeated Experiments).

Suggested Ridership Data Collection Guidelines

Based on the findings from the literature, the survey of Virginia transit agencies, and the analysis of ridership data, the following general data collection guidelines are proposed.

- *Transit agencies should be aware that following the NTD sampling template does not always guarantee that the error in ridership estimation will be small.* If resources are available, increasing the number of sampled weeks should always be considered. This is because the sample size is affected by the coefficient of variation of the input data (e.g., historical records from the previous year). If the coefficient of variation of the input data is small, there will be risk of underestimating the needed sample size for a target year with a different demand pattern than occurred in the previous year. Alternatively, a larger factor A (e.g., using 1.5 instead of 1.25 in the original sampling procedure) for the margin of safety in the sample size determination equation (5) should be considered to help reduce estimation errors.

$$n_{target} = \left(\frac{Z_{0.95}}{\text{Precision}} \times \frac{std_{input}}{\mu_{input}} \right)^2 \times A = \left(\frac{1.96}{0.1} \times CV_{input} \right)^2 \times A \qquad (5)$$

- *When sampling approaches are used, transit agencies should report detailed information on how the sample size was determined, regardless of whether sampling was done at trip level, route level, or system level.* This could be assisted with the inclusion of an example to illustrate the adopted procedure to determine the minimum sample size.

- *Transit agencies gathering APC data for reporting should always verify and adjust the collected data based on comparisons with benchmark data (e.g., manual ride checks) before reporting.* Validating and certifying APCs can help minimize inaccuracy of the machines. However, this does not guarantee that collected data will be of high quality. Necessary manual correction done alongside ride checks should be considered to adjust the potential inflation of counts due to drivers' own on-and-off actions as well as non-riders entering and exiting with inquiries. It is especially suggested to exclude such counts for low-ridership routes. The adjustment/correction information should be documented and reported along with the estimated ridership.

- *Transit agencies that rely on driver counting should be aware of the potential human errors and discrepancies between drivers.* Verification by independent ride checkers can be considered, and driver training should always include a component on the best practices of data collection while on duty. Agencies should not simply assume a constant error rate among drivers for data correction.

- *When farebox data are used for reporting, a ridership estimation procedure should be established to account for factors such as the use of coupon books, passes, or discounts for certain riders.* The complexity of fare collection systems (e.g., discounted fares and fare-free services) among transit agencies makes it difficult to standardize farebox-based ridership estimation. Each agency's established estimation procedure should be transparent and be reported along with the estimated ridership.

- *If full data were collected, agencies should track the time-series data and regularly review abnormal values to verify if they are accurate outliers (e.g., winter weather lowering a day's ridership) or inaccurate (e.g., far beyond the bus capacity) and requiring correction.* An agency's ridership estimation procedure should include clearly defined thresholds for flagging abnormal records, e.g., data points more than two standard deviations away from the mean, data points 50% more or less than the average for the previous day(s) or the same period for previous weeks, etc. As an example, those circled points in Figure 20 may deserve special attention.

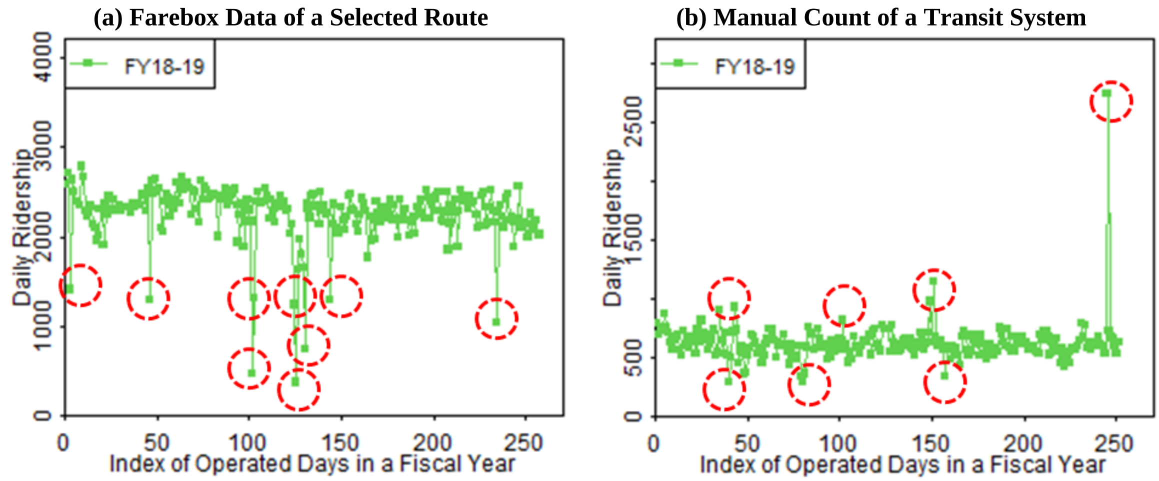


Figure 20. Example of Tracking Time Series Records for Screening Abnormal Data.

- *When there are systematic changes in data collection techniques during the reporting period, the changes should be documented and reported along with the estimated ridership.* Some agencies have mentioned that their data were not comparable across years, even for the same types of data collection approaches (e.g., because of changing the vendors of APCs). After any systematic changes in data collection techniques, the data quality needs to be re-assessed, and a note on the updated data quality going forward should be reported.

- *When multiple types of ridership data are available, agencies should compare their quality and report the most reliable one.* Some agencies may have multiple types of data, such as from fareboxes, APCs, and manual counting with mobile data terminals or clickers. Data points with notable inconsistencies among different data sources deserve special attention. For example, in Figure 21, the highlighted values for each route deserve special attention.

ROUTE	1		2		3		4		5		6	
Date	Farebox	APC	Farebox	APC	Farebox	APC	Farebox	APC	Farebox	APC	Farebox	APC
7/12/2018	211	203	183	200	208	217	111	141	144	117	82	112
7/13/2018	234	243	150	188	182	208	77	96	181	200	83	109
7/14/2018	23	196	119	134	9	131	0	94	108	128	51	67
7/16/2018	0	241	170	185	0	226	0	117	217	259	0	113
7/17/2018	0	233	0	191	0	214	0	136	0	203	0	124
7/18/2018	166	189	142	181	219	237	88	121	176	194	119	122
7/19/2018	211	252	150	184	203	227	102	135	156	211	71	110
7/20/2018	246	300	141	162	169	226	99	119	179	207	138	144

Figure 21. Example of Crosschecking Different Types of Data.

- *When daily ridership records are transferred and stored, the raw data should be reviewed and corrected in a timely way.* For example, there may be typos or machine failures. Such correction efforts were mentioned by many agencies, but no detailed information was available. Any correction procedure applied should be documented and reported.

- *Ridership data should be stored in electronic files that are convenient to use.* For example, saving data in a commonly used format (e.g., .csv file) makes it more accessible than file formats that are difficult to manipulate or proprietary (e.g., PDF file and scheduling software, respectively).

- *If sampling approaches are considered for estimating ridership, the sampled counts should cover different periods.* As some routes may have seasonal patterns, a segmented sampling procedure should be considered. For example, instead of randomly sampling the days of a year, organizing the sampling procedure based on weeks in different months can better capture the seasonal changes in demand. This will also be more practical in terms of managing the data collection process than randomly sampling individual days.

DISCUSSION

This study was not focused on conducting field observational tests to evaluate the quality of different ridership data collection solutions. Thus, it does not generalize regarding the exact accuracy of each solution. In fact, based on discussions with representatives from different transit agencies in Virginia, the errors associated with each type of data collection approach can be affected by many factors such as unexpected non-passenger interference (e.g., bus drivers causing overcounting in APCs when they get on and off the bus for breaks) and miscounting by drivers using mobile data terminals. The heterogeneity of these factors often makes it difficult to systematically correct the potential errors in data. This suggests a need to direct some efforts to the training of raw data collectors, including bus operators, and to establish formal data review and quality control procedures at each agency. Some agencies have indicated that they have some internal actions to check their data, but they were often case by case. Establishing a data quality control procedure can help mitigate more obvious issues due to human errors and bias due to technology limits.

The research team examined sampling issues based on daily data from transit agencies in Virginia. Although both system-level and route-level data were analyzed, this study did not test the NTD sampling approach based on trip-level data, as the historical trip-level data were not available. Nevertheless, trip-level data are expected to have large variations among different trips and routes, and if demand is not stable from year to year, the sample size may be underestimated for a target year.

It should be noted that some agencies are looking at becoming fare-free for a term or indefinitely (GRTC, 2021; DRPT, 2021). As a result, collecting ridership data based on fareboxes may be challenging. Instead of relying on fareboxes, these agencies should consider alternative solutions such as APCs and manual counting for obtaining ridership data.

CONCLUSIONS

- *Unified agreement is lacking in both the literature and among transit agencies in Virginia on the performance of different ridership data collection approaches due to diverse approaches in terms of data sampling, estimation, evaluation metrics, and comparison pairs.* Few studies were found that addressed the methodological issues of estimating ridership based on sampled data.

- *There are diverse ridership data collection technologies employed by the transit agencies in Virginia.* These data collection technologies include APCs, fareboxes, manual survey, mobile apps, AFC, and so on. Although APC counts and electronic farebox data were the two major sources, neither was used by more than one-third of the responding agencies to obtain ridership data. Considering the common availability of electronic fareboxes on transit vehicles, agencies that have not leveraged such data can consider using it as a valuable source for estimating ridership.

- *Route-level data are available at many transit agencies in Virginia.* The most frequently collected level of ridership data was the route level for fixed-route buses, as more than 50% of the surveyed agencies indicated that they maintained route-level data.

- *Regardless of the raw data type, some agencies considered scaling methods for obtaining long-term ridership estimates.* The sampled data were often scaled to estimate annual ridership. Nevertheless, for those agencies reporting to NTD, about two-thirds did not estimate ridership, but rather directly summed and reported the actual daily counts.

- *Farebox data and APC data from the same vehicle were found to be inconsistent, and their differences also varied among different routes.* Reporting ridership data based on raw farebox or APC data makes it difficult for DRPT to judge the data quality and how it differs among transit agencies.

- *If sampled data were used for estimating ridership data, a sample size determined using historical data cannot guarantee the accuracy of estimated ridership in a target year.* An increased sample size will help reduce estimation errors. In addition, considering segmented sampling will help account for seasonal variations and make the sampling implementation and management easier.

- *A set of ridership data collection guidelines has been suggested*. The developed guidelines can assist transit agencies in improving their ridership data quality and allow DRPT to judge ridership data quality and differences among transit agencies.

RECOMMENDATIONS

1. *DRPT should provide the guidelines developed as a part of this study or a modified version of them to transit agencies to facilitate their ridership data collection practices.* Many surveyed transit agencies have expressed their interest in learning the findings of this study.

Distributing the guidelines will help facilitate information-sharing about the study's findings and may help transit agencies improve their own data collection practices. DRPT should periodically (e.g., every 3-5 years) revisit the guidelines to ensure they remain current and relevant, as there could be improvements or notable changes in ridership data collection practices.

2. *DRPT should require the submission of ridership data collection methods and correction (adjustment) procedures used for each mode by each transit agency along with its final reported ridership data.* These supporting documents will provide more transparent information on how the reported ridership data by mode are developed, efforts made to improve data quality, and the potential issues present; all of which can help DRPT better understand and defend the quality of the reported data.

IMPLEMENTATION AND BENEFITS

Researchers and the technical review panel (listed in the Acknowledgements) for the project collaborate to craft a plan to implement the study recommendations and to determine the benefits of doing so. This is to ensure that the implementation plan is developed and approved with the participation and support of those involved with VDOT and DRPT operations. The implementation plan and the accompanying benefits are provided here.

Implementation

Following the publication of this report, DRPT will incorporate these recommendations into its ridership data reporting process by Winter 2022-2023. DRPT's Statewide Transit Planning Manager will facilitate the distribution of the developed guidelines with this project report to transit agencies in Virginia. DRPT will periodically (e.g., every 3-5 years) review the developed guidelines to account for possible changes or improvements in ridership data collection practices.

Benefits

The benefit of implementing Recommendation 1 will be improved information-sharing among transit agencies in Virginia. Transit agencies will be able to learn about practices of peer agencies. If agencies improve the quality of their ridership data by applying data collection practices described in the guidelines, they will benefit from better data for their own planning and operational decision-making.

The benefit of implementing Recommendation 2 will be improved quality of ridership data through enhanced data collection and processing approaches. In addition, the quality of the reported ridership data and potential issues associated with the reported data will be clearer to DRPT. This will facilitate improved decision-making for planning and funding at the state level (including project prioritization).

ACKNOWLEDGMENTS

The research team highly appreciates the valuable comments and suggestions from the TRP members, including Tiffany Dubinsky (Project Champion and Statewide Transit Planning Manager; DRPT), Shane Sawyer (Multimodal Planning Programs Manager, VDOT TMPD), Cheryl Spain (Program Support Specialist Senior, City of Harrisonburg), and Hyun Cho (Research Scientist, VTRC). The authors also thank the following individuals for their assistance: Amy O'Leary and Peter Ohlms from VTRC, all survey respondents from transit agencies, all transit agency staffers who shared ridership data, and staff at the Old Dominion University Social Science Research Center for helping conduct the survey.

REFERENCES

Boyle, D., 2008. *TCRP Synthesis 77: Passenger Counting Systems - A Synthesis of Transit Practice.* Transportation Research Board of the National Academy of Sciences, Washington, D.C.

Brakewood, C.E., 2014. *Quantifying the Impact of Real-time Information on Transit Ridership.* Georgia Institute of Technology.

Cambridge Systematics, 2014. *VDOT Travel Demand Modeling Policies and Procedures.* VDOT Project ID: 30681-03-02. Transportation and Mobility Planning Division, Virginia Department of Transportation (VDOT).

Chu, X., 2018. *Development of an NTD Tool for Vanpool Services*. Final Report. Center for Urban Transportation Research, University of South Florida, Tampa.

DART, 2018. *Quarterly Operating, Financial Performance, and Compliance Report*. First Quarter FY 2018. Dallas Area Rapid Transit. Accessed April 23, 2020.

DRPT, 2019. *DRPT Statewide Integrated Mobility Initiative*. Virginia Department of Rail and Public Transportation, Virginia.

DRPT, 2021. *Transit Ridership Incentive Program (TRIP) Application Guidelines*. Fiscal Year 2022 Mid Grant Cycle Application. Virginia Department of Rail and Public Transportation, Virginia.

Dunlap, M., Li, Z., Henrickson, K., Wang, Y., 2016. Estimation of Origin and Destination Information from Bluetooth and Wi-Fi Sensing for Transit. *Transportation Research Record 2595* (1), 11-17.

Fehr & Peers, 2018. *Josephine County TNC First Last Mile Service Solutions: Ridership and Cost Estimates.* Final Report Combined Appendices. Fehr & Peers.

FCDOT, 2015. *Connections 2015 Fairfax County Comprehensive Transit Plan.* Fairfax Connector Ridership Survey. Fairfax County Department of Transportation (FCDOT).

Furth, P.G., 2000. *TCRP Synthesis 34: Data Analysis for Bus Planning and Monitoring - A Synthesis of Transit Practice.* Transportation Research Board of the National Academy of Sciences, Washington, D.C.

García-Bunster, G., Torres-Torriti, M., 2008. Effective Pedestrian Detection and Counting at Bus Stops. In: *Proceedings of the 2008 IEEE Latin American Robotic Symposium*, 158-163.

GRTC, 2021. DRPT Awards $8m State Grant to GRTC to Study Zero Fares. http://ridegrtc.com/news-initiatives/press-releases/drpt-awards-8m-state-grant-to-grtc-to-study-zero-fares. Accessed January 7, 2022.

HRT, 2018. *Transit Development Plan FY2018-FY2027.* Hampton Roads Transit. Hampton, Virginia.

Jung, Y.-J., Casello, J.M., 2019. Assessment of the Transit Ridership Prediction Errors Using AVL/APC Data. *Transportation 47*, 2731–2755.

Kimley Horn, IBI Group, 2019. *DRPT Statewide Integrated Mobility Initiative.* Final Report. Virginia Department of Rail and Public Transportation, Virginia.

Kostakos, V., Camacho, T., Mantero, C., 2013. Towards Proximity-based Passenger Sensing on Public Transport Buses. *Personal and Ubiquitous Computing 17* (8), 1807-1816.

Kotz, A.J., Kittelson, D.B., Northrop, W.F., 2015. Novel Vehicle Mass-based Automated Passenger Counter for Transit Applications. *Transportation Research Record 2563* (1), 37-43.

Lawson, C.T., Muro, A., Krans, E., 2021. Forecasting Bus Ridership Using a "Blended Approach". *Transportation 48,* 617-641.

Lu, A., Reddy, A., 2012. Strategic Look at Friday Exceptions in Weekday Schedules for Urban Transit: Improving Service, Capturing Leisure Markets, and Achieving Cost Savings by Mining Data on Automated Fare Collection Ridership. *Transportation Research Record 2274* (1), 30-51.

Monast, K., Vaughan, C.L., Jones, P., Carnes, C., 2017. *Using Security Cameras to Count Transit Passengers.* Research and Analysis Group, North Carolina Department of Transportation.

Oberli, C., Torres-Torriti, M., Landau, D., 2010. Performance Evaluation of UHF RFID Technologies for Real-time Passenger Recognition in Intelligent Public Transportation Systems. *IEEE Transactions On Intelligent Transportation Systems 11* (3), 748-753.

Peterson, D., 2013. *Effortless Passenger Identification System.* Final Report for Transit IDEA Project 70. Transportation Research Board, Washington, DC.

Reddy, A., Lu, A., Kumar, S., Bashmakov, V., Rudenko, S., 2009. Entry-only Automated Fare-Collection System Data Used to Infer Ridership, Rider Destinations, Unlinked Trips, and Passenger Miles. *Transportation Research Record 2110* (1), 128-136.

RTAMS, 2020. *Ridership Statistics*. http://www.rtams.org/rtams/ridershipHome.jsp. Accessed May 7, 2020.

Saavedra, M.R., 2010. *An Automated Quality Assurance Procedure for Archived Transit Data from APC and AVL Systems.* University of Waterloo.

SFMTA, 2020. *Municipal Transportation Quality Review - Fiscal Years 2016-2017 & 2017-2018.* San Francisco Municipal Transportation Agency.

Shireman, M.T., 2011. *Using Automatically Collected Data for Bus Service and Operations Planning.* Massachusetts Institute of Technology.

Sound Transit, 2015. *Transit Ridership Forecasting Methodology Report*. Sound Transit, Seattle, Washington.

Strathman, J.G., Kimpel, T.J., Callas, S., 2005. Validation and Sampling of Automatic Rail Passenger Counters for National Transit Database and Internal Reporting at TriMet. *Transportation Research Record 1927* (1), 217-222.

Tétreault, P.R., El-Geneidy, A.M., 2010. Estimating Bus Run Times for New Limited-stop Service Using Archived AVL and APC Data. *Transportation Research Part A: Policy and Practice 44* (6), 390-402.

Toon, B., 2018. *ABQ RIDE Ridership Statistics by Route.* Fiscal Year 2018 (July 2017 through June 2018). https://www.cabq.gov/transit/documents/fy18-annual-productivity-summary.pdf. Accessed May 13, 2020.

VDOT, 2019. *SMART SCALE*. http://www.vasmartscale.org. Accessed September 5, 2020.

WAVE Transit, 2018. *Wave Transit's Short-term Efficiencies and Long-term Governing Model*. Final Report. Wave Transit Authority, City of Wilmington, North Carolina.

Yahiaoui, T., Khoudour, L., Meurie, C., 2010. Real-time Passenger Counting in Buses Using Dense Stereovision. *Journal of Electronic Imaging 19* (3), 031202.

Yang, S., Wu, Y.-J., Marion, B., Moses, I.E., 2015. Identification of Transit Farebox Data Errors: Impacts on Transit Planning. *Public Transport 7* (3), 457-473.

Yang, T., Zhang, Y., Shao, D., Li, Y., 2010. Clustering Method for Counting Passengers Getting in a Bus with Single Camera. *Optical Engineering 49* (3), 037203.

APPENDIX A – SURVEY SAMPLE

Dear Transit Project Manager/Officer,

A research team at Old Dominion University (ODU) is leading a research project to examine current practices of transit ridership data collection among transit agencies. This project is for the Virginia Department of Transportation (VDOT) and the Department of Rail and Public Transportation (DRPT). The project is particularly interested in learning your TYPICAL practices for transit ridership data collection and data reporting and any lessons your agency has learned in the long term. It should be noted that if COVID-19 pandemic has affected your long-term ridership data collection practices, we would like to learn that as well.

Below is a link to a survey to collect information regarding the practices in your organization. You can use a computer or a smart phone to open the link and complete the survey. It should be noted that it might be easier to view and complete the survey on a computer. Your participation is voluntary, and your responses are confidential. The survey responses will only be analyzed and reported in an aggregated way. The survey will take about 15~25 minutes to complete.

We would value survey responses from both agencies that are existing National Transit Database reporters and agencies that are non-reporters. In addition, agencies providing different types of transit services (e.g., fixed routes, on-demand, vanpools, etc.) are all invited to take the survey. Please click on the link below to complete the survey. Please help complete the survey by **November 13th, 2020**.

Survey link: {Survey link was added here}

If some other staff in your organization are managing different types for ridership data, please help share the survey to them. Thank you very much.

If you have any questions about the survey, please contact Dr. Hong Yang (hyang@odu.edu) or Tancy Vandecar-Burdin (tvandeca@odu.edu). Our DRPT point of contact for this project is Tiffany Dubinsky (tiffany.dubinsky@drpt.virginia.gov). Our VDOT point of contact is Peter Ohlms (peter.ohlms@vdot.virginia.gov). If you need a printed copy of the survey, please also let us know. We would also greatly appreciate it if you could also share the survey with related peers within your agency and those at other Virginia transit agencies / organizations that collect ridership data.

Your participation and responses are greatly appreciated.

Best Regards,

ODU Research Team for VDOT & DRPT Transit Ridership Data Project

Questionnaire for Transit Ridership Data Collection Practices

> *This survey is about the current practices in transit ridership data collection among Virginia transit agencies. The survey is part of an ongoing project led by a research team at Old Dominion University (ODU) with support from the Virginia Department of Transportation (VDOT) and the Department of Rail and Public Transportation (DRPT). The Principal Investigators of the project are Drs. Hong Yang, Sherif Ishak, and Kun Xie at ODU. The information collected will help researchers understand and learn the best practices in collecting ridership data among different transit agencies. The survey will take about 15~25 minutes to complete.*
>
> *If you have any questions about the survey, please contact Dr. Hong Yang (hyang@odu.edu) or Tancy Vandecar-Burdin (tvandeca@odu.edu). If you need a printed copy of the survey, please also let us know. We would greatly appreciate it if you could also share the survey with* peers at other Virginia transit agencies / organizations that collect ridership data. *Thank you very much.*

(multiple ☐ / single choices ○)

1. Agency/organization Name: Click or tap here to enter text.

2. Contact Person's Information. The research team may reach you for clarifying any uncleared comments or questions only. The contact information will not be used in any other cases irrelevant to this project.
   a. Name: Click or tap here to enter text.
   b. Work Phone: Click or tap here to enter text.
   c. Work Email: Click or tap here to enter text.

3. Approximate number of vehicles in fleet in your agency/organization: Click or tap here to enter text.
4. What type(s) of area(s) is the agency serving? (check all that apply)
   a. Urban ☐
   b. Suburban ☐
   c. Rural ☐

5. At what levels are ridership data collected for each of the following modes? (Segments are between different stops; trips are between certain origins and destinations; routes are fixed bus routes.) (check all that apply)

Service/vehicle Type	Stop	Segment	Trip	Route	System	Other
Bus	☐	☐	☐	☐	☐	☐
Commuter bus	☐	☐	☐	☐	☐	☐
Bus rapid transit	☐	☐	☐	☐	☐	☐
Trolley-style bus	☐	☐	☐	☐	☐	☐
Vanpool	☐	☐	☐	☐	☐	☐

Paratransit	☐	☐	☐	☐	☐	☐
Other vehicle type(s) Click or tap here to enter text.	☐	☐	☐	☐	☐	☐

Note: Services such as heavy rail, light rail, and commuter rail are not considered in this survey.

6. Please indicate how frequently the data are accessible to your staff for each of the following items. (check all that apply)

Data	Daily	Weekly	Monthly	Quarterly	Annually	As needed
System ridership	☐	☐	☐	☐	☐	☐
Route-level ridership	☐	☐	☐	☐	☐	☐
Route segment ridership	☐	☐	☐	☐	☐	☐
Stop-level boarding/alighting	☐	☐	☐	☐	☐	☐

7. Please indicate how frequently the data are shared outside your agency (e.g., with Department of Rail and Public Transportation (DRPT), National Transit Database (NTD), etc.) for each of the following items. (check all that apply)

Data	Daily	Weekly	Monthly	Quarterly	Annually	As needed
System ridership	☐	☐	☐	☐	☐	☐
Route-level ridership	☐	☐	☐	☐	☐	☐
Route segment ridership	☐	☐	☐	☐	☐	☐
Stop-level boarding/alighting	☐	☐	☐	☐	☐	☐

8. What tools does your agency use to collect ridership data? (check all that apply)
   a. Automated passenger counters (APC) ☐
   b. Electronic fareboxes ☐
   c. Automated fare collection (AFC) devices ☐
   d. Manual survey (by staff other than drivers) ☐
   e. Mobile app ☐
   f. Manually reviewing videos recorded by in-vehicle cameras ☐
   g. Other tools (please explain) Click or tap here to enter text.

   Note: Tools such as pencil & paper and mechanical clickers used by drivers will be considered as other tools.

9. In what formats are ridership data stored? (check all that apply)
   a. Text file (e.g., .txt file, Word, or Google Docs) ☐
   b. Spreadsheet (e.g., Excel or Google Sheets) ☐
   c. Relational database (e.g., Oracle) ☐
   d. Specialized software ☐
   e. Handwritten ledger ☐

f. Other format (please explain) Click or tap here to enter text.

10. Are the ridership data considered publicly available data (i.e., freely available)?
    a. Yes ○
    b. No ○

11. Are software tools used to analyze collected ridership data?
    a. Yes (please provide the name of the software tools)
       Click or tap here to enter text.
    b. No ○

12. What data processing steps are applied to validate collected ridership data? (check all that apply)
    a. Compare with fare revenue ☐
    b. Look for unexplained variations across trips ☐
    c. Compare totals across days ☐
    d. Rely on the professional judgment of analysts ☐
    e. Use an automated program to analyze data ☐
    f. Compare boarding vs. alighting totals ☐
    g. Compare with manual counts ☐
    h. Other data processing step (please explain) Click or tap here to enter text.

13. How satisfied is your organization with the quality of ridership data obtained using from the following data sources?

Data Source	Very Satisfied	Satisfied	Neutral	Unsatisfied	Very Unsatisfied	Not Available
Automated Passenger Counter (APC)	○	○	○	○	○	○
Electronic Farebox	○	○	○	○	○	○
Automated Fare Collection (AFC)	○	○	○	○	○	○
Manual Survey	○	○	○	○	○	○
Mobile App	○	○	○	○	○	○

Click or tap here to enter text.

14. What are your agency's primary purposes for collecting ridership data? (check all that apply)
    a. Identify least and most productive routes ☐
    b. Identify candidate stops for elimination/addition ☐

c. Compile National Transit Database (NTD) reports ☐
d. Compile reports for DRPT ☐
e. Determine maximum passenger loads only in COVID-19 period ☐
f. Determine maximum passenger loads in normal times ☐
g. Monitor schedule adherence and/or running times ☐
h. Adjust schedules (add/delete trips, change headways) ☐
i. Validate travel demand models ☐
j. Transit service planning for transit-oriented development ☐
k. Calculate other performance measures ☐
l. Other purposes (please explain) Click or tap here to enter text.

15. How are the raw ridership data primarily sampled?
    a. Based on sampled stops ○
    b. Based on sampled routes ○
    c. Based on all stops /routes ○
    d. Other sampling method (please explain) Click or tap here to enter text.

16. How are the raw ridership data transferred from data collection devices/tools to storage? (check all that apply)
    a. Direct downlink with a physical connection (e.g., Ethernet connection to a computer and download data ☐
    b. Retrieval at garage without a physical connection (e.g., wirelessly downloading data at garage via vendor software applications) ☐
    c. Real-time retrieval or periodic remote retrieval with software applications ☐
    d. Removable storage medium (i.e., memory stick, memory card) ☐
    e. Manual data entry ☐
    f. Unknown ☐
    g. Other Click or tap here to enter text.

17. Which supplemental details about ridership are collected in addition to counts? (check all that apply)
    a. None ☐
    b. Timestamps ☐
    c. GPS coordinates ☐
    d. Fare types ☐
    e. Transfer status ☐
    f. Special rider types ☐
    g. Other (please explain) Click or tap here to enter text.

18. What percentage of the fleet for the major service type (e.g., fixed-route bus) in your organization is equipped with automated passenger counters (APCs)?
    a. None ○
    b. 1-25% ○
    c. 26-50% ○
    d. 51-75% ○
    e. 76-100% ○

19. For the portion of the fleet equipped with automated passenger counters (APCs), how are the APC-equipped vehicles assigned to routes? (check all that apply)
    a. No APC-equipped vehicles in my organization ☐
    b. All vehicles are equipped with APCs ☐
    c. Rotation of equipped vehicles between routes periodically ☐
    d. Dedicated equipped vehicles to selected routes ☐
    e. Other (please explain) Click or tap here to enter text.

20. Which types of automated passenger counter (APC) technology are used? (check all that apply)
    a. None ☐
    b. Based on infrared light ☐
    c. Based on Bluetooth/Wi-Fi ☐
    d. Based on video ☐
    e. Other (please explain) Click or tap here to enter text.

21. What percentage of the fleet (i.e., owned and subcontracted) is equipped with electronic fareboxes?
    a. None ○
    b. 1-25% ○
    c. 26-50% ○
    d. 51-75% ○
    e. 76-100% ○

22. For the portion of the fleet equipped with electronic fareboxes, how are the farebox-equipped vehicles assigned to routes? (check all that apply)
    a. No electronic farebox-equipped vehicles in my organization ☐
    b. All vehicles are equipped with electronic fareboxes ☐
    c. Rotation between routes periodically ☐
    d. Dedicated to selected routes ☐
    e. Other (please explain) Click or tap here to enter text.

23. What percentage of the fleet (i.e., owned and subcontracted) is equipped with automated fare collection (AFC) devices?

a. None ○
b. 1-25% ○
c. 26-50% ○
d. 51-75% ○
e. 76-100% ○

24. For the portion of the fleet equipped with automated fare collection (AFC) devices, how are the AFC-equipped vehicles assigned to routes? (check all that apply)
    a. No AFC-equipped vehicles in my organization ☐
    b. All vehicles are equipped with AFC devices ☐
    c. Rotation between routes periodically ☐
    d. Dedicated to selected routes ☐
    e. Other (please explain) Click or tap here to enter text.

25. How often does your organization conduct manual surveys to collect ridership data?
    a. Never ○
    b. Weekly ○
    c. Monthly ○
    d. Quarterly ○
    e. Annually ○
    f. Other (please explain) Click or tap here to enter text.

26. At which locations are manual surveys conducted? (check all that apply)
    a. None ☐
    b. At stops/ or specific sites ☐
    c. Onboard vehicles ☐
    d. Review video from onboard camera systems ☐
    e. Other (please explain) Click or tap here to enter text.

27. On what proportion of routes do you conduct manual surveys?
    a. None ○
    b. 1-25% ○
    c. 26-50% ○
    d. 51-75% ○
    e. 76-100% ○

28. For each type of raw data collection method your agency uses, please indicate if you have used a sampling technique (e.g., collected data from a number of sampled routes/bus stops/time periods). For those unused raw data sources, just leave the lines unchecked.

Raw Data Source for Ridership Estimation	Sampling Technique Involved
APC passenger counts	○ Yes; ○ No; ○ We do not have such data.

Electronic Farebox passenger counts	○ Yes; ○ No; ○ We do not have such data.
Manual passenger counts	○ Yes; ○ No; ○ We do not have such data.
Mobile ticketing passenger counts	○ Yes; ○ No; ○ We do not have such data.
AFC revenue data	○ Yes; ○ No; ○ We do not have such data.
Farebox revenue data	○ Yes; ○ No; ○ We do not have such data.
Mobile app's revenue data	○ Yes; ○ No; ○ We do not have such data.
Order data (e.g., reservation records of paratransit)	○ Yes; ○ No; ○ We do not have such data.
Others Click or tap here to enter text.	○ Yes; ○ No;

29. Based on the collected raw data, please indicate the methods used to obtain longer-term ridership estimates. (check all that apply)
    - *Scaling method: e.g., one day count multiplied by 7 to get the weekly ridership*
    - *Regression method: ridership is predicted by regression equation approaches*
    - *Weighted average: weighted sum of different data sources*

Raw Data Source	Scaling Method	Regression Method	Weighted Average	Other Methods
APC passenger counts	☐	☐	☐	Click or tap here to enter text.
Electronic Farebox passenger counts	☐	☐	☐	Click or tap here to enter text.
Manual passenger counts	☐	☐	☐	Click or tap here to enter text.
Mobile ticketing passenger counts	☐	☐	☐	Click or tap here to enter text.
AFC revenue data	☐	☐	☐	Click or tap here to enter text.
Farebox revenue data	☐	☐	☐	Click or tap here to enter text.
Mobile app's revenue data	☐	☐	☐	Click or tap here to enter text.
Order data (e.g., reservation records of paratransit)	☐	☐	☐	Click or tap here to enter text.

Others Click or tap here to enter text.	☐	☐	☐	Click or tap here to enter text.

30. If multiple data sources are used for your primary service (e.g., fixed-route service or paratransit service), how do you primarily integrate the data for longer term ridership estimation?
    a. Not applicable ○
    b. Data imputing (e.g., use historical data to infer missing counts) ○
    c. Weighted average of different data ○
    d. Other (please explain) Click or tap here to enter text.

31. Which type of National Transit Database (NTD) report does your agency employ?
    a. None ○
    b. Full data report (monthly and annually) ○
    c. Reduced data report (annually) ○

32. Please select National Transit Database (NTD) ridership-related data measurements tracked by your agency. (check all that apply)
    a. None ☐
    b. Unlinked passenger trips (UPT) ☐
    c. Linked passenger trips (LPT) ☐
    d. Total distance traveled by all passengers (PMT) ☐
    e. Other (please explain) Click or tap here to enter text.

33. Please select the NTD procedures your agency employs. (check all that apply)
    a. None ☐
    b. Count all passengers without sampling ☐
    c. Use the NTD-provided sampling algorithms ☐
    d. Use a non-NTD sampling procedure with a qualified statistician ☐
    e. Other (please explain) Click or tap here to enter text.

34. If your organization reports ridership data to the National Transit Database (NTD), please describe any steps that are taken to validate ridership data for NTD reporting purposes. If not, please put "not reporting." Click or tap here to enter text.

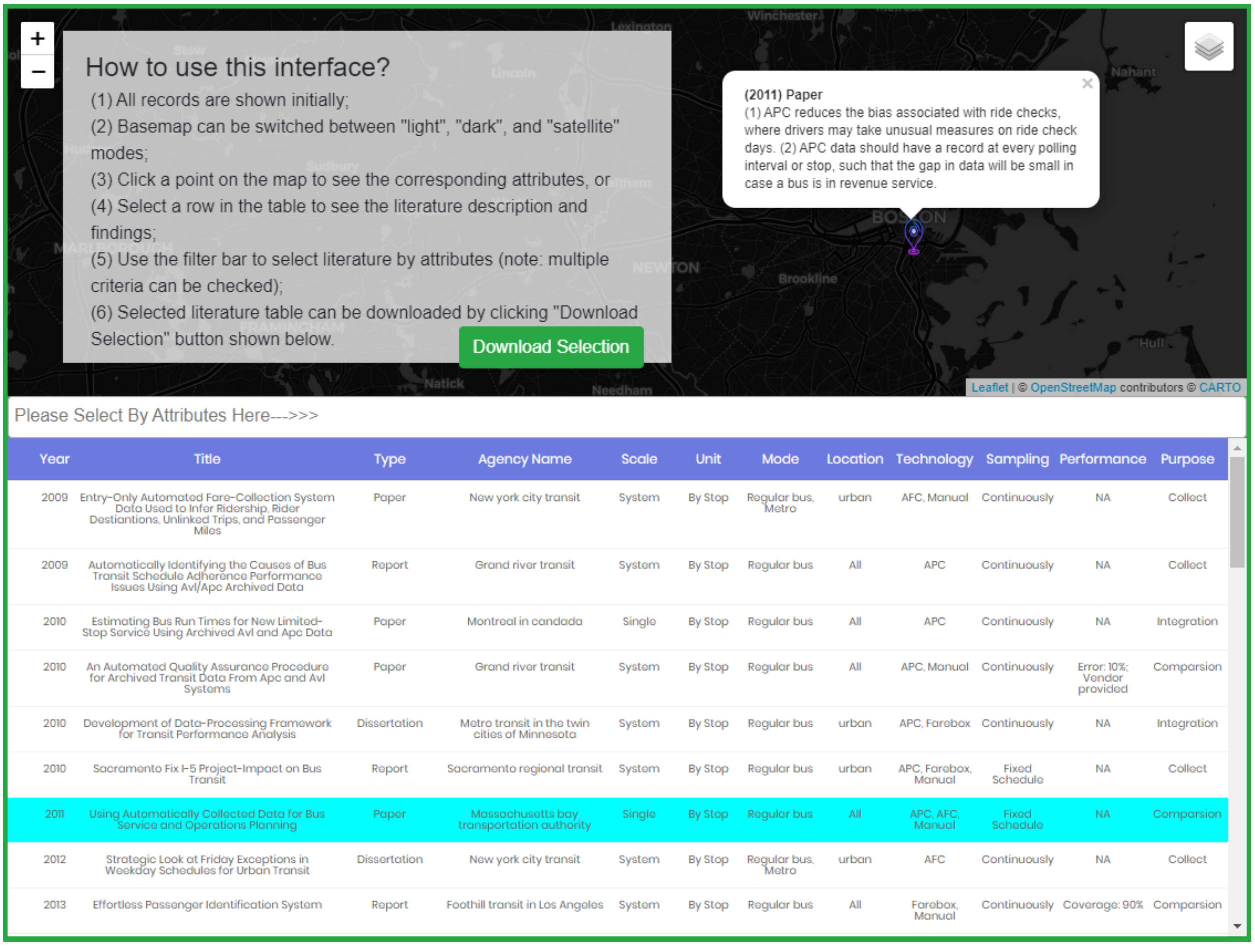
35. If the ridership data collection process(es) have changed since 2018, please identify the applicable improvements. (check all that apply)
    a. No changes in recent years ☐
    b. Expanded electronic data collection effort ☐
    c. Expanded data collection scale ☐
    d. Improved ridership data estimation approach ☐
    e. Other (please explain) Click or tap here to enter text.

36. Before the COVID-19 pandemic, if your agency planned to change its ridership data collection process(es) by 2022, please indicate how.
    a. No plans to change ☐
    b. Plan to expand electronic data collection effort ☐
    c. Plan to expand data collection scale ☐
    d. Plan to improve ridership data estimation approach ☐
    e. Other (please explain) Click or tap here to enter text.

37. Has your agency changed / does your agency plan to change its ridership data collection process(es) in responding to the impact of COVID-19 pandemic?
    a. Yes (please explain) ○
       Click or tap here to enter text.
    b. No ○

Thank you very much for your participation! If you have any questions, please feel free to contact the research team: Dr. Hong Yang (hyang@odu.edu) or Tancy Vandecar-Burdin (tvandeca@odu.edu).

APPENDIX B – SCREENSHOT OF THE DEVELOPED LITTERATURE REVIEW INTERFACE



- Web Interface for Summarized Literature: http://senselane.com/pubtransit/

APPENDIX C – LIST OF AGENCIES THAT RESPONDED TO SURVEY

Completed

1. Bay Aging, dba Bay Transit
2. Fairfax Connector
3. Winchester Transit
4. Radford Transit: operated by New River Valley Community Services
5. Greater Lynchburg Transit Company (GLTC)
6. Central Shenandoah Planning District Commission/BRITE
7. Greater Roanoke Transit Company
8. Virginia Department of Rail and Public Transportation
9. Williamsburg Area Transit Authority
10. Blacksburg Transit
11. City of Suffolk - Suffolk Transit
12. Loudoun County Transit
13. DASH (Alexandria Transit Company)
14. Charlottesville Area Transit
15. Potomac and Rappahannock Transportation Commission-OmniRide
16. Arlington County
17. Four County Transit
18. Hampton Roads Transit
19. RideFinders
20. Fredericksburg Regional Transit
21. RideShare/TJPDC
22. Virginia Regional Transit
23. Pony Express - Town of Chincoteague

24. Jaunt, Inc.

25. Pulaski Area Transit

26. Town of Bluefield/Graham Transit

27. City of Harrisonburg Department of Public Transportation

Incomplete for some questions:

28. Danville Transit System

29. Petersburg Area Transit (PAT)

30. Fredericksburg Regional Transit

31. City of Bristol Virginia

32. Mountain Empire Older Citizens, Inc. Transit Department

33. Hampton Roads Transportation Planning Organization

34-39. Six Unknown Agencies (Name not provided)

APPENDIX D – OTHER SAMPLING METHODS USED BY DATA TYPE

This appendix lists respondents' verbatim answers in the "other" category for methods related to each data type.

Responses for other methods – APC passenger counts:

Actual data collected daily
N/A (4)
None (2)
We count all passengers
We do not estimate future ridership

Responses for other methods – Electronic Farebox passenger counts:

N/A (2)
None
Not used
We count all passengers
We do not estimate future ridership
We don't have this
We use 100 % count

Responses for other methods – Manual passenger counts:

Actual data collected daily
done on vanpools
Driver hand collectors
N/A
None - we don't sample this data.
See previous question "Other" answer
Trend analysis
We count all passengers
We do not estimate future ridership

Responses for other methods – Mobile ticketing passenger counts:

Do not have mobile ticketing
N/A (5)
None
None - we don't sample this data.
Not used
We count all passengers
We do not estimate future ridership
We don't have this

Responses for other methods – AFC revenue data:

N/A (5)
None
None - we don’t sample this data.
Not used
We count all passengers
We do not estimate future ridership
We don't have this

Responses for other methods – Farebox revenue data:

Actual data collected daily
Manual count twice weekly
N/A
None - we don’t sample this data.
Not used
We count all passengers
We do not estimate future ridership

Responses for other methods – Mobile app’s revenue data:

Do not have mobile ticketing
N/A (5)
None
None - we don’t sample this data.
Not used
We count all passengers
We do not estimate future ridership
We don't have this

Responses for other methods – Order/reservations data:

Actual data collected daily
NA (2)
None
None - we don’t sample this data.
Trend analysis
We count all passengers
We do not estimate future ridership
We use 100 % count

Responses for other methods – Other raw data sources:

N/A (4)
None

None - we don't sample this data.
Not used
We count all passengers
We do not estimate future ridership

APPENDIX E – STEPS TAKEN TO VALIDATE RIDERSHIP DATA FOR NTD REPORTING PURPOSES

This appendix lists respondents' verbatim answers to the prompt "If your organization reports ridership data to the National Transit Database (NTD), please describe any steps that are taken to validate ridership data for NTD reporting purposes."

Compared to monthly data submitted to DRPT

Conduct reasonableness checks such as passengers per hour, and compare to previous years, considering changes in service level and other external factors.

Daily and monthly reports are compared to previous trends looking on a monthly level. If necessary, reports can be reviewed for previous trends at route and service day level.

Daily validation of data compared to trends and data point outliers

Data collected and reported by DRPT

Enterprise reports all vanpool data for our region

[Agency] validates ridership value by internal staff review of the data.

In vanpooling we rely on the data provided to us from the vanpool vendors and vanpool coordinators

Internal review and validation of data.

Manual Count

Not reporting

Review ridership sources

Ridership data validated based on previous year performance

Ridership is recorded and compared with daily operations reports.

The UPTs come straight out of the farebox system for bus

Validated through the RouteMatch Paratransit and Fixed Route Scheduling and Dispatch Software.

Validation occurs by ticket sales and passenger "check in" procedures.

Validation step including daily, weekly, monthly and annually internal reporting based on manual data collection compared to APC derived counts.

[Agency] performs manual comparisons to certify the accuracy of its APC system. All data is reviewed and compiled into a report from a qualified statistician. Then we go through the process of submitting paperwork to NTD for certification of the APC system.

We count all passengers, our contractor sends us copies of the Passenger sheets the operators use and we compare those to the report that send us as the end of the month.

We hire a consultant firm through NVTC that does this work.

We use farebox data for NTD ridership reporting. We are working to validate against available APC data, but we also use professional judgment to identify any data that does not look correct. Within the next two years, we hope to have more APC's and obtain NTD certification to use APCs.